2015

DnaG as an antibacterial target

Zorik Chilingaryan
University of Wollongong

UNIVERSITY OF WOLLONGONG

School of Chemistry

DnaG as an Antibacterial Target

Zorik Chilingaryan

"This thesis is presented as part of the requirement for the Award of the Degree of Doctor of Philosophy of the University of Wollongong"

March 2015

ABBREVIATIONS

ADME	Absorption, distribution, metabolism and elimination
COOT	Crystallographic object-oriented tool kit
CPMG	Car-Purcell-Meiboom-Gill
CSP	Chemical shift perturbation
DEAE	Dimethylaminoethyl
DMSO	Dimethyl sulfoxide
DnaGC	C-terminal domain of DnaG primase
DQF-COSY	Double quantum filter correlation spectroscopy
DTT	Dithiothreitol
EDTA	2, 2', 2", 2'"-(ethane-1,2-diyldinitrilo)tetraacetic acid
ESI-MS	Electrospray ionization-mass spectroscopy
ExoI	Exonuclease I
FBDD	Fragment-based drug design
FBS	Fragment-based screening
FP	fluorescence polarization
FPLC	Fast protein liquid chromatography
fs	Femtosecond
GE	Group efficiency
h	hour
HBD	Helicase-binding domain
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HMBC	Heteronuclear multiple bond correlation
HSQC	Heteronuclear single quantum coherence
HTS	High throughput screening
ITC	Isothermal titration calorimetry
kDa	Kilodaltons
LB	Lysogeny broth
LE	Ligand efficiency
MD	Molecular dynamics
MES	2-(*N*-morpholino)ethanesulfonic acid
MQW	Milli-Q water
MRSA	Methicillin-resistant *Staphylacoccus aureus*

MW	Molecular weight
NMR	Nuclear magnetic resonance
NOESY	Nuclear Overhauser enhancement spectroscopy
ns	Nano second
OB fold	Oligonucleotide/oligosaccharide-binding fold
OD_{600}	Optical density at 600 nm wavelength
PCS	Peudocontact shift
PDB	Protein data bank
PEG	Polyethylene glycol
PPI	Protein-protein interactions
PRE	Paramagnetic relaxation enhancement
ps	Pico second
*R*g	Radius of gyration
RCD	RNA polymerase and C-terminal domains of DnaG
RMSD	Root mean square deviation
rRMSD	residue-based RMSD
RPD	RNA polymerase domain of DnaG
RU	Response units
S/N	Signal to noise ratio
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SPR	Surface plasmon resonance
SSB	Single stranded DNA-binding protein
SSB-Ct	C-terminus peptide of SSB
STD	Saturation transfer difference
Tris	Tris(hydroxymethyl)aminomethane/Trizma ™
VMD	Visual molecular dynamics
waterLOGSY	Water-ligand observed via gradient spectroscopy
ZBD	Zinc binding domain

ABSTRACT

The rise of "superbugs" – antibiotic resistant infection-causing bacteria - poses a catastrophic threat to public health, and prompts the investigation of new drug targets. In this work, the inhibition of specific proteins in bacterial DNA replication machinery – a rich source of new targets - was investigated.

The single stranded DNA-binding protein, SSB, is an interaction hub that engages in vital interactions with several partner proteins through a flexible C-terminal peptide motif (DDDIPF; SSB-Ct). Based on available crystal structures of SSB-Ct with partner proteins, molecular dynamics simulations were used to identify mobile elements of SSB-Ct and important interactions in SSB-Ct binding-pockets.

Fragment-based screening using various methods was undertaken, targeting the SSB-binding partner DnaG, the replicative primase from *E. coli*. Initially two fragment libraries (by Zenobia Therapeutics and the Monash Institute of Pharmaceutical Science) containing ~1600 fragments were screened against DnaG primase using SPR competition and STD-NMR assays. The binding confirmation and binding site location of identified fragments was confirmed by ligand- and protein-based NMR techniques. Common features were identified in some fragment hits: carboxylic group containing aromatic compounds.

In silico fragment-to-hit optimisation led to the identification of phenoxy-phenyl tetrazoles; ZINC72447025 and CDS001350 with 1.3 mM and 1.2 mM binding affinities. It was hypothesized that these compounds could inhibit other proteins that bind to SSB-Ct. The tetrazole derivatives and other fragment hits were shown binding to other SSB-Ct binding partners such as *E. coli* χ subunit of Pol III, PriA, RNAse HI and *A. baumannii* χ.

This work demonstrates the potential for inhibitions SSB/binding partner interactions, and the possibility of inhibiting multiple SSB-binding proteins with similar ligands.

ACKNOWLEDGEMENTS

I would like to thank my supervisor Associate Professor Aaron Oakley for his wisdom, guidance, patience and constant encouragement throughout this long and ultimately rewarding endeavor. Thanks for teaching me crystallography even without having any crystals. Special thank you to Professor Nick Dixon for sharing his knowledge of DNA replication, problem solving and suggestion paramagnetic NMR experiments.

Thanks to Dr. Allen Lo, who was very supportive from the beginning of my PhD, helping me with every experiment in the beginning to become more independent, Dr Slobodan Jergic for his input to my project. Special thanks to Dr. Robyn Mansfield for innumerable thoughtful discussions on many aspects of NMR, big help to learn not simple field of protein NMR, Dr. Ann Kwan (USyd) who selflessly gave her time helping to work with NMR, set up any experiments that I could come up without ever asking for credit. Dr. Steve Heady and A/Prof. Martin Scanlon (MIPS) for giving insides on many aspects of NMR in drug design. Professor Gottfried Otting (ANU) for opening new horizons to paramagnetic NMR to me. Yao and Tiffany for warm welcoming in the lab. Amy and Jacob for a great company and atmosphere in the office, and proofreading the thesis and the rest of Oakley and Dixon group.

Lastly, I am very grateful to my parents who have supported me during this long rigorous journey far away from home. Very special thank you to Sergey, whose friendship and support I had throughout these years.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

.

Chapter 1

General Introduction

1.1 The Discovery of Antibiotics: an Overview

In everyday life, the human body interacts with an untold number of microorganisms. While most are harmless, some are dangerous enough to be deadly. Many antibiotics available today were originally extracted from natural products. After the discovery of the first antibiotics in early 1930s, many more effective compounds have been introduced into everyday clinical practice. **Figure 1.1** provides a historical overview of antibiotic discovery.

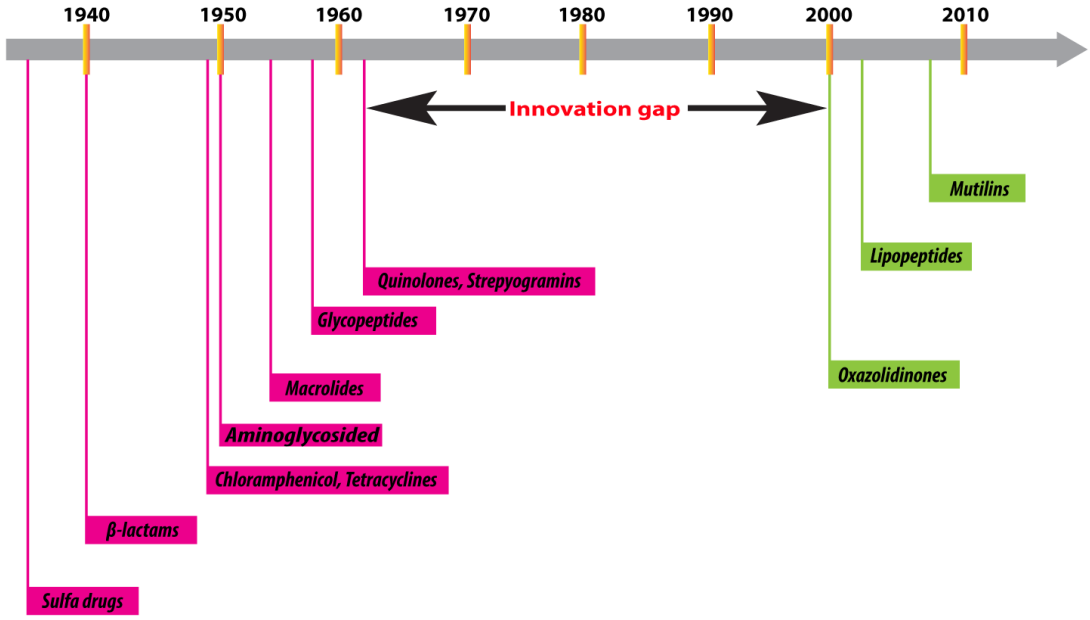


Figure 1.1. Timeline of discovery of antibiotics. Based on (Fischbach and Walsh, 2009).

In recent decades, only a few new groups of antibiotics have been identified; from 1960 till 2000 there was an "innovation gap" where no new classes of antibiotics were introduced (J H Powers, 2004). In fact, only a handful of new classes of antibacterial drugs have been approved by FDA since 2000: mutilins, lipopeptides, and oxazolidones.

Pathogenic bacteria resistant to antibiotics have become a matter of grave concern. Bacterial pathogens resistant to all clinically available antibiotics have been identified (Heinemann *et al.*, 2000; Longley and Johnston, 2005). Strains in question include methicillin-resistant *Staphylococcus aureus* and various strains of *Mycobacterium tuberculosis*, *Acinetobacter baumannii* and *Pseudomonas aeruginosa* (**Figure 1.2**).

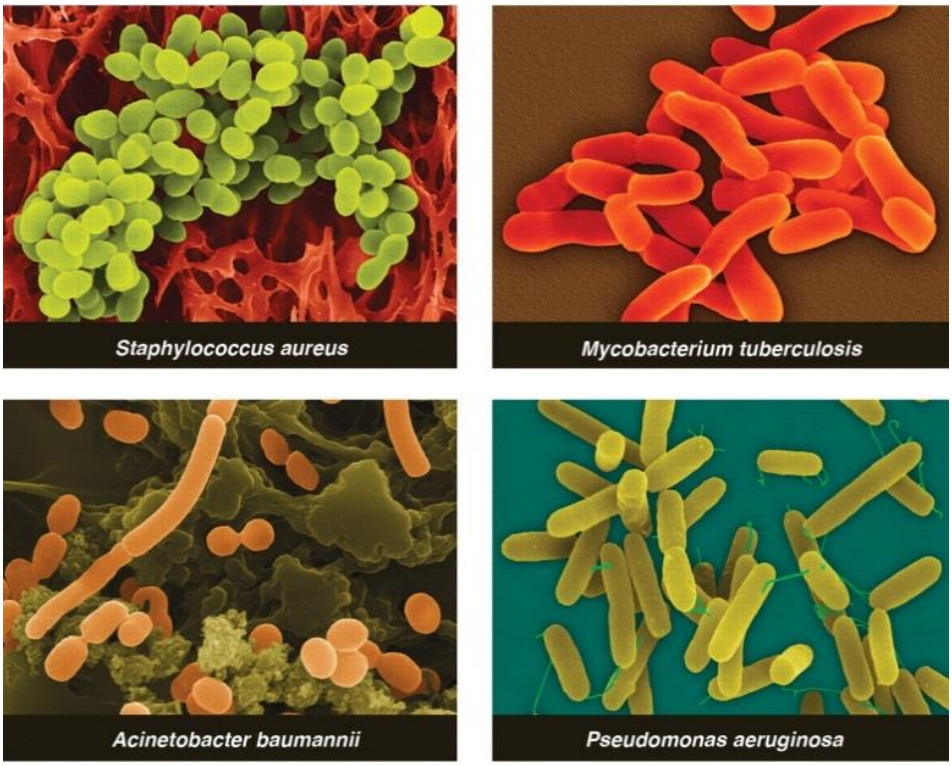


Figure 1.2. Micrographs of the resistant strains of *S. aureus*, *M. tuberculosis*, *A. baumannii*, *P. aeruginosa*. Figure reproduced from (Fischbach and Walsh, 2009).

1.2 Steps Involved in the Discovery of New Drugs

The discovery of new drugs takes place in several steps and involves a range of specialists from biologists to clinicians. Today, drug discovery tends to begin with the identification of targets followed by screening of chemical libraries to identify compounds inhibiting or activating the function of the target. The identified compounds are optimised with respect to increasing the affinity and specificity to the target. The optimised compound is then tested for ADME (absorption, distribution, metabolism and elimination), followed by intensive preclinical and clinical trials, where the drug candidate is tested on humans. If these are successful, the new drug is registered with regulatory bodies (**Figure 1.3**). It is important to note that failure may be encountered at any step.

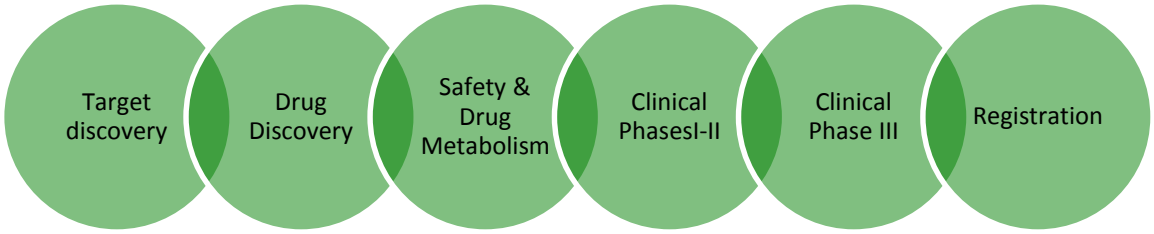


Figure 1.3. Steps involved in the drug discovery process starting from target identification to registration of drug by a regulatory organization (in the US: the Food and Drug Administration).

1.3 Drug Targets

Target identification is one of the most important steps in drug discovery projects. Identification of a suitable target increases the chance of success for further hit-to-lead identification. A recent analysis of AstraZeneca's research and development programs has revealed that the failure of discovery projects is very often related to a lack of information regarding the biology of a particular disease, which commonly is attributable to a poorly chosen target (Cook *et al.*, 2014). The importance of target validation should therefore never be underestimated.

In recent years, an increasingly popular drug discovery strategy is "multi-target drug discovery". The identification of drugs affecting multiple targets increases the chances of success, and overcomes the problem of resistance. Several agents like multikinase and EGFR family inhibitors have been identified that modulate multiple targets (Morphy, 2010).

1.3.1 Antibiotic Targets

Since 1930 a few groups of antibiotics have been identified targeting cell wall synthesis, protein synthesis etc. All clinically used antibiotics and their targets are summarized in **Figure 1.4** (Lewis, 2013).

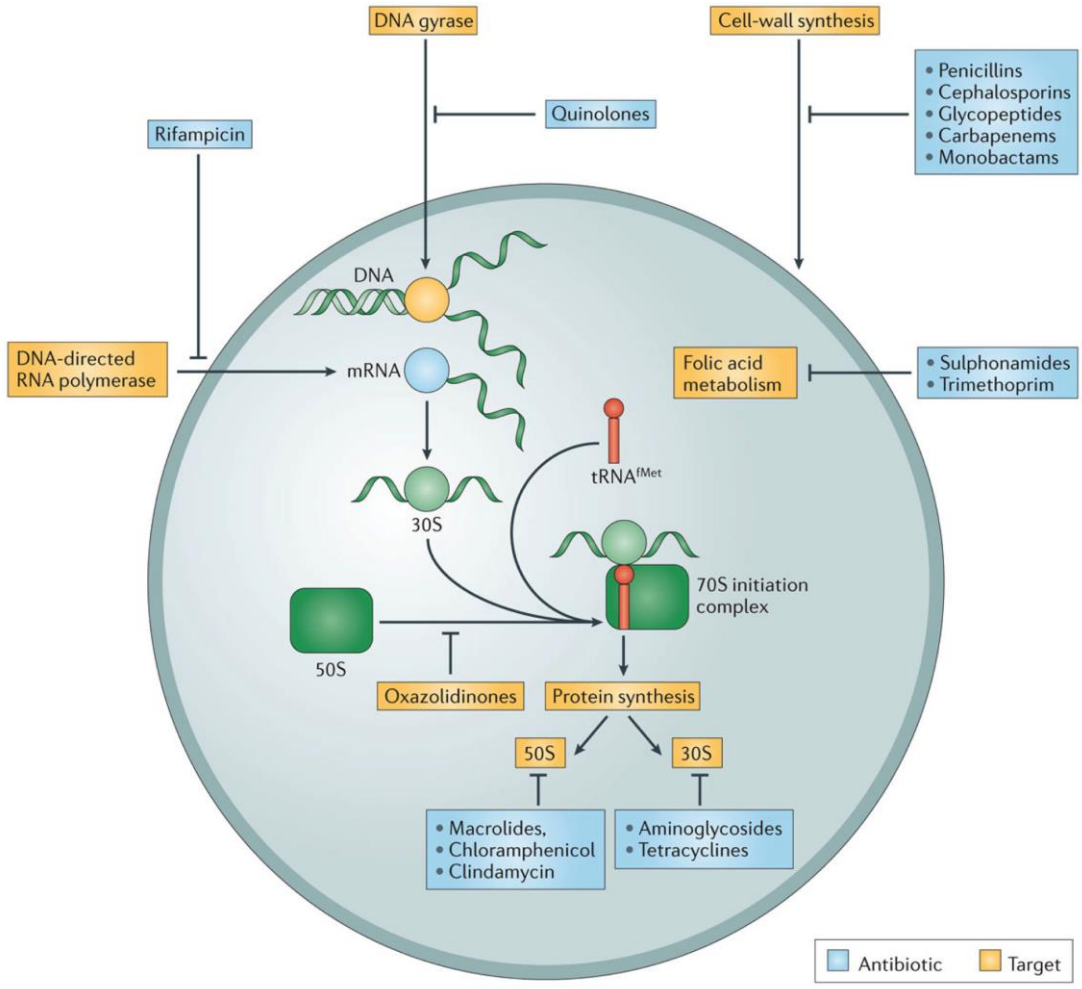


Figure 1.4. A bacterial cell with all currently available processes targeted in antibiotic discovery. Figure reproduced from (Lewis, 2013).

In order to overcome antibacterial resistance, there is a need to identify new drug targets. Bacterial DNA replication, recombination and repair is an underexploited drug target, involving many protein-protein and protein-DNA interactions (Schaeffer *et al.*, 2005; Robinson *et al.*, 2012).

1.4.1 High-Throughput Screening Versus Fragment-Based Screening

Large pharmaceutical companies traditionally relied on high-throughput screening (HTS) to identify hits. HTS involves the use of sizeable chemical libraries (10^3 to 10^7 compounds), comprised of chemically diverse, largely drug-like structures (Walters and Namchuk, 2003; Macarron *et al.*, 2011; Phillip Gribbon and Andreas, 2005). HTS has been reliant on the use of automated enzymatic or cell-based assays for screening, which are prone to false positive hits due to non-specific and off-target interactions

(Philip Gribbon and Sewing, 2003). One relatively new but increasingly popular approach in drug discovery is fragment-based screening (FBS), an alternative to HTS with a broader choice of screening method (Rees *et al.*, 2004). FBS starts from relatively small building blocks known as *fragments*. The fragments are typically chosen to satisfy the '*Rule of Three*', introduced by Astex pharmaceuticals (Congreve *et al.*, 2003). The rule of three prescribes that fragments should have a molecular mass of < 300 Da, up to three hydrogen bonds donors, up to three hydrogen bond acceptors, number of rotatable bonds < 3 and calculated logP (clogP) of < 3. These rules help to ensure that lead compounds with drug-like physicochemical properties are ultimately produced. Fragments generally contain aromatic rings with chemically diverse functional groups. During optimisation of a fragment hit, the MW is increased, along with the potency of the compound. This stepwise engineering of a small molecule minimizes unfavourable steric and electrostatic interactions in the targeted pocket (Congreve *et al.*, 2003). Additional rules have recently been proposed regarding the various physicochemical properties of drug-like compounds, particularly regarding their oral bioavailability (Veber *et al.*, 2002). Others have proposed alternative rules such as those related to the permeability of the blood-brain barrier (Oprea, 2002; Andrew M Davis *et al.*, 2001; Pardridge, 1995).

The ultimate goal of FBS is to construct inhibitors with K_Ds in the nanomolar range from fragment hits. Although fragments tend to bind with K_Ds in the millimolar range, they tend to make 'high quality' i.e. highly favourable-interactions compared with hits from HTS (**Figure 1.5, 1.6**).

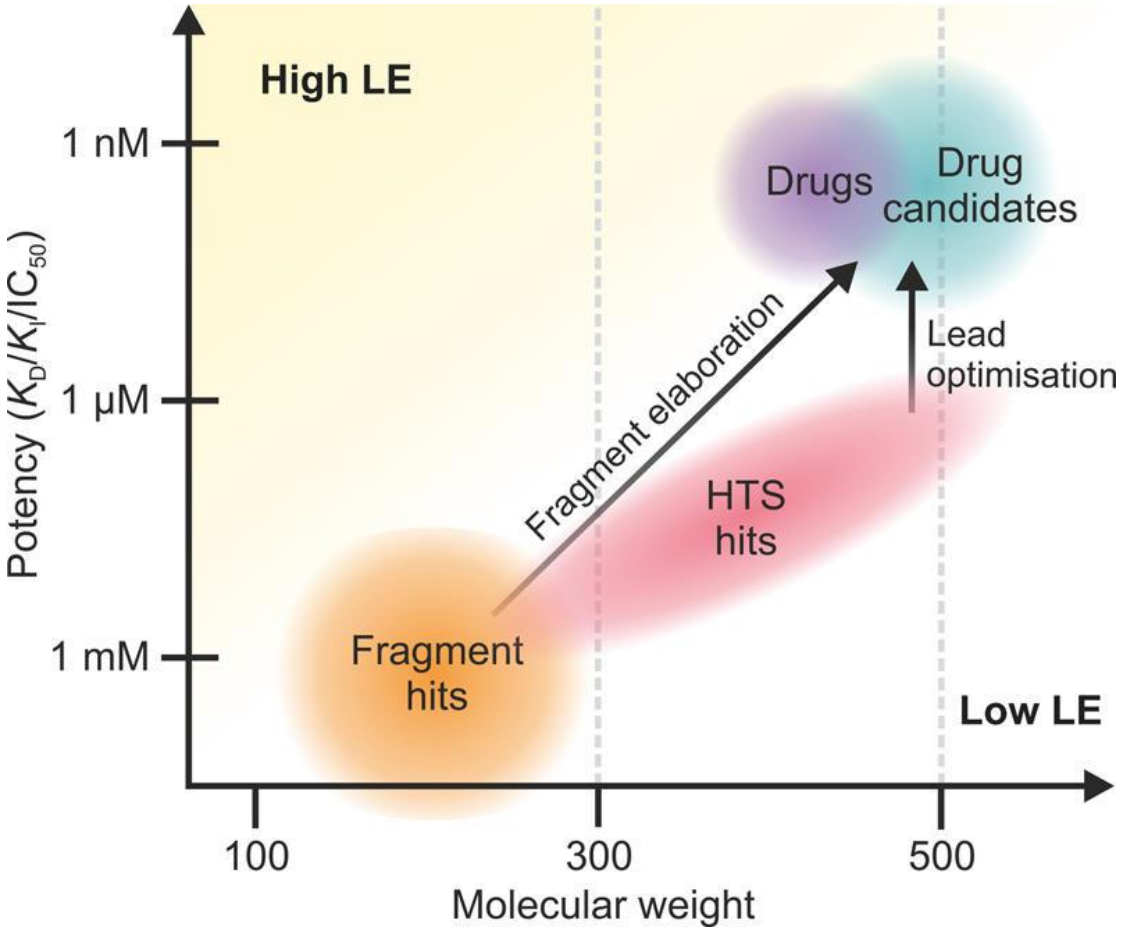


Figure 1.5. The dependency of molecular weight on potency; The improvement of potency is highlighted in relation to LE. Reproduced from (Scott et al., 2012).

In order to estimate the 'drug-likeness' of hit molecules and derivatives, several metrics are used. Ligand efficiency (LE) and group efficiency (GE) are two examples of such metrics (Hopkins *et al.*, 2004; Kuntz *et al.*, 1999; Verdonk and Rees, 2008). LE is defined as -ΔG/HAC, where ΔG is the Gibbs free energy of binding, and HAC is the heavy (non-hydrogen) atom count. In contrast, GE stresses the relative importance of functional groups either added to or excluded from a hit molecule during optimisation. It is expressed as GE = - $\Delta\Delta G_b/\Delta HA$: where $\Delta\Delta G_b = \Delta G_b(B)-\Delta G_b(A)$ and ΔHAC =HAC(B)-HAC(A), where A and B are the existing and optimised molecules respectively, and $\Delta G_b(B)$ and $\Delta G_b(A)$ are the free energies of binding of existing and optimised molecules (Verdonk and Rees, 2008). All of the aforementioned requirements directly or indirectly influence each other emphasizing the complexity of the fragment-to-lead optimisation.

When available, multiple fragment hits in different parts of a target site can be linked to build lead compounds with higher affinity (**Figure 1.6**).

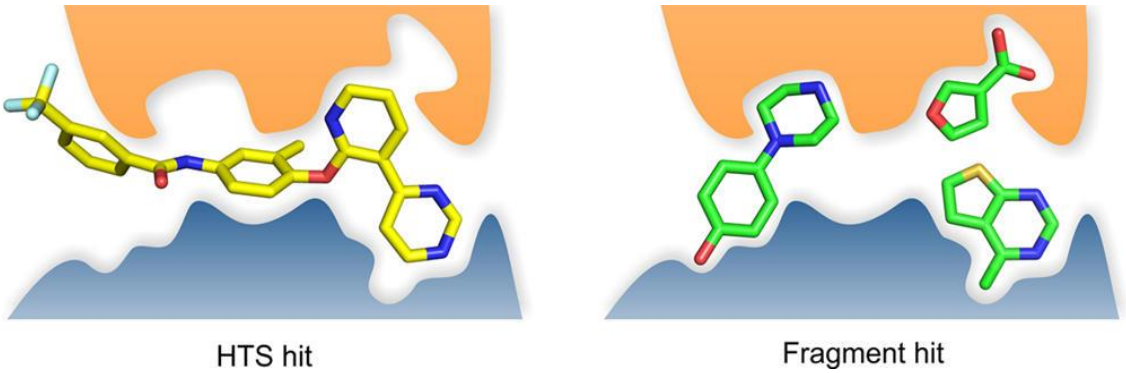


Figure 1.6. A comparison of HTS and FBS that visualizes the predicted ligand fitting into the binding pocket of a target. Figure reproduced from (Scott et al., 2012).

Although HTS has been used successfully against many targets, the hit rate is often very low. Moreover, the optimisation of large drug-like molecules is challenging. There is a correlation between molecule size and poor pharmacokinetic properties. With FBS, low affinity hits are extended gradually, with physical properties carefully controlled, in order to increase their affinity to the target. The comparison of HTS to FBS is summarized in **Table 1.1** (Rees *et al.*, 2004).

Table 1.1. A comparison of HTS and FBS showing the two main screening approaches. Reproduced from (Rees et al., 2004).

	HTS	**FBS**
Emphasis on	potency	efficiency
MW range	~250 – 600	~150 - 300
Hit activity range (K_D)	~30 μM - nM	~mM – 30 μM

Fragment library can be screened against the target using various biophysical techniques (X-ray, NMR, SPR) in order to detect non-covalent biding, followed by fragment elaboration, during which the lead compound will be generated. The entire process is summarized in **Figure 1.7** (Scott *et al.*, 2012).

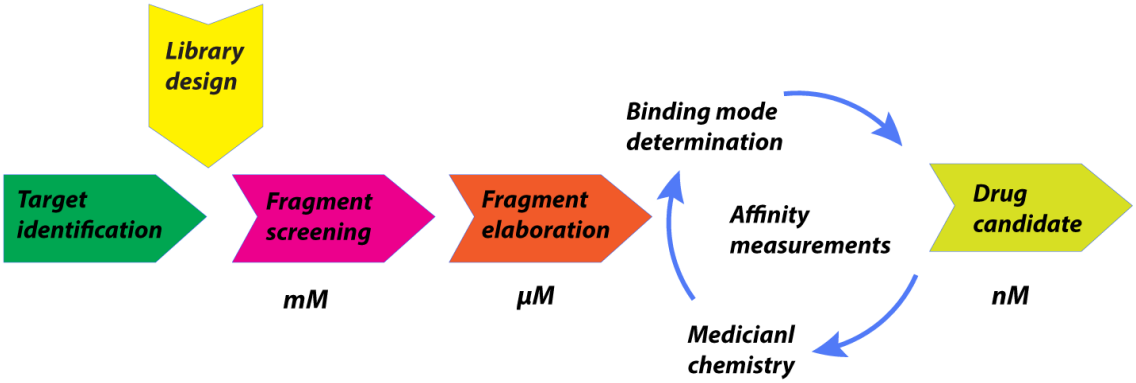


Figure 1.7. Highlighting comparative screening methods, showing the fragment-based drug discovery process. Based on (Scott et al., 2012).

1.4.2 A Fragment-based Screening Methods

The choice of screening method for FBS is crucial. A "three-stage biophysical cascade" has been proposed as a good fragment screening platform, starting with a preliminary screen (thermal shift assay), validation (ligand binding observed by NMR) and structural characterisation (X-ray crystallography) (Mashalidis *et al.*, 2013). The use of different biophysical methods at different stages of fragment screening helps to validate the results and eliminate false positives. At present there are only a few methods that can detect weak protein-fragment and nucleic acid-fragment interactions, e.g., NMR, X-ray crystallography, surface plasmon resonance (SPR), fluorescence polarization (FP), isothermal titration calorimetry (ITC) and more recently mass spectroscopy and thermal shift assays. A comparison of some of these methods is given in **Table 1.2** below. Each of the approaches mentioned has an associated affinity detection range. FBS typically begins with small molecules often with K_D in the millimolar range (**Figure 1.8**). Any method used to screen fragments should be sensitive enough to detect these interactions.

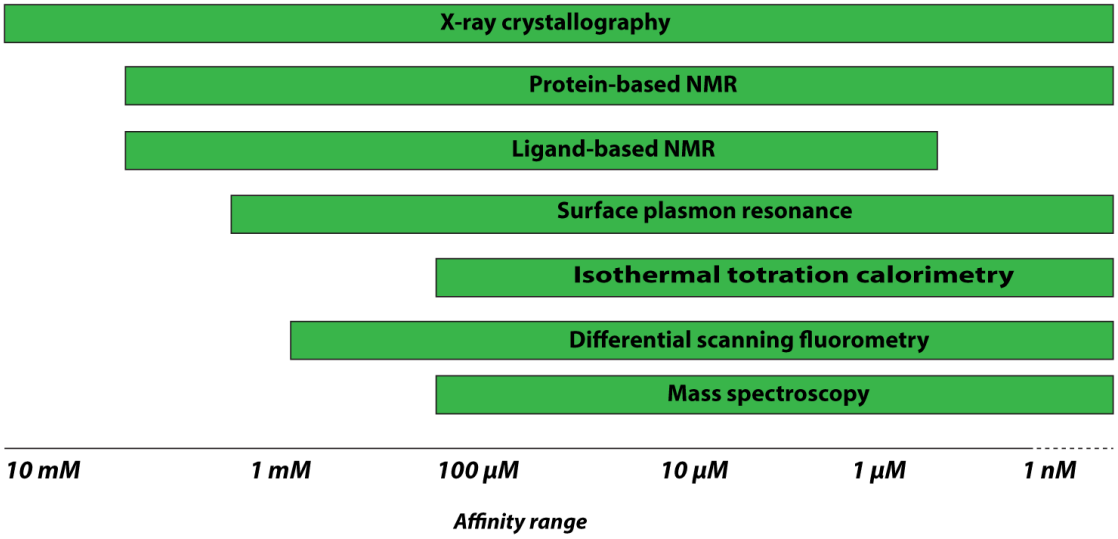


Figure 1.8. Representation of the affinity range detected by various biophysical methods. Based on (Hubbard and Murray, 2011).

Subsequent *in silico* screening tools, such as virtual screening, molecular docking, and molecular dynamics (MD) simulations are commonly used to extend our knowledge beyond the available spectroscopic limits.

Table 1.2. Comparison of screening methods. Table reproduced with some modifications from (Dalvit and Stockman, 2002).

	X-ray crystallography	**NMR**		**SPR**	**ITC**	**FP**
		Ligand-observed	*Protein-observed*			
Response dependence	-	[EL]	[EL]/[E_T]	[EL]/[E_T]	[EL]/[E_T]	[EL]/[E_T]
Data acquisition Speed	Slow	Fast	Fast	Fast	Slow	Fast
Protein size sensitivity	No	No	Yes	Yes	No	Yes
Protein concentration	μM to mM	Low μM	μM	μM	μM	Low μM
Fragment concentration	μM to mM	μM	μM to mM	μM to mM	μM to mM	μM to mM
Binding site on target	Yes	No	Yes	No	No	No
Binding epitope on ligand	Yes	Yes	No	No	No	No
Dynamic range	-	Large	Medium-large	Small	Small	Small
Intrinsic sensitivity	-	Low	Low	High	High	High

1.4.2.1 X-ray Crystallography

The information provided by X-ray crystallography is in the form of atomic three-dimensional structures of target biomolecules and complexes with ligands. The use of crystallization robots, synchrotron beam-lines, robotic sample changers, and automated fitting of electron density by programs such as AutoSolve by Astex Therapeutics (Mooij *et al.*, 2006) makes crystallography a useful high throughput method for solving the structures and observing the orientation of ligand in the binding pocket. Required for FBS by X-ray crystallography are robust crystals amenable to soaking experiments. Another essential requirement is that in general, the resolution should be greater than 2.5 Å so that ligand-binding poses can be interpreted unambiguously. Also, the target biding site should be unhindered by crystal contacts (Warren *et al.*, 2012). Crystalizing the target can be time consuming and challenging, depending on the protein. Crystallography occasionally yields false positives that can sometimes be explained by the use of DMSO as a solvent for fragments, and by artefacts arising from crystallization conditions (such as a conformation of macromolecule dominant in the crystal form but not highly populated in solution). By X-ray crystallography, only one single conformation of the ligand and/or macromolecule is typically observed. If diffraction quality is sufficiently high, it is possible to detect multiple conformers of protein side chains and/or fragments. The availability of crystal structures of target proteins makes it possible to apply computational methods such as MD simulations to study the dynamic behaviour of the complex structure.

1.4.2.2 Nuclear Magnetic Resonance

Nuclear Magnetic Resonance (NMR) is a method extensively used in structural and chemical biology. Methodological advances have made it possible to apply NMR to molecules with various NMR-active nuclei (e.g. ^{1}H, ^{13}C, ^{15}N, ^{31}P, ^{19}F) to study structures, function and interactions (Grzesiek and Sass, 2009). In 2000, scientists at Abbott Laboratories used NMR as a fragment screening tool for the first time (Hajduk, 2006). Improvements in technique and increased automation has allowed the detection of ligand binding *in vivo* (Betz *et al.*, 2006). From a drug discovery perspective, NMR experiments can be divided into two groups: protein-based (HSQC, HMQC, TOCSY, NOESY etc.) and ligand-based methods (STD, CPMG, WaterLOGSY etc.) (Dalvit and Stockman, 2002). Ligand-based NMR methods were introduced about 15 years ago

(Dalvit *et al.*, 2001; Mayer and Meyer, 1999; Hajduk *et al.*, 1997). They are now widely used as a primary screening steps. These methods rely on the observation of NMR signals from ligands. Their advantage is that there is no effective upper limit on the size of the protein. NMR-based methods can even been applied to screening for ligands that bind to membrane proteins (Meinecke and Meyer, 2001). Ligand-based methods are more efficient for larger proteins, which have longer rotational correlation times (τ_c) and exhibit a negative Nuclear Overhauser Effect (NOE) that improves spin diffusion and therefore saturation transfer from protein to bound ligand. Furthermore, protein consumption is low compare to other biophysical techniques (Dalvit, 2009; Mayer and Meyer, 1999). Cocktails containing up to 10 fragments each can be screened at once. A limitation of ligand-based methods is that no structural data is obtained for the binding location of ligands on the biomolecule. However, it is capable of providing ligand epitope (Mayer and Meyer, 2001). A useful property of the ligand-based methods is the ability to detecting weak binding, i.e., K_D values of up to 10 mM (Meyer and Peters, 2003). For both STD and waterLOGSY to be successful, the off-rate has to be fast in relation to the relaxation time. The corresponds to dissociation constants in the mM to μM range. The smallest dissociation constant reported for waterLOGSY was ~ 0.1 μM (Dalvit *et al.*, 2001).

Another ligand-based method is Carr-Purcell-Meiboom-Gill (CPMG) NMR. This method exploits differences in relaxation properties between small molecules and proteins, and is based on measurements of relaxation properties of small molecules that differ in the free and bound states. It has the ability to detect tighter binders with K_D values in the nM range (Dalvit and Stockman, 2002).

In the absence of protein crystals, ligand-based methods may be followed by ^{1}H-^{15}N and/or ^{1}H-^{13}C HSQC experiments in order to identify the residues involved in binding. ^{1}H-^{15}N HSQC is highly sensitive to weak interactions with K_D as high as 10 mM (Zuiderweg, 2002). These protein-based NMR experiments require isotopically labelled protein. Despite the limitations noted below, protein-based NMR is useful for proteins of MW up to 30-35 kDa. Recent developments have made possible the use of NMR for larger proteins and membrane proteins with size of up to 1,000,000 Da using TROSY-based approaches (Zhu and Yao, 2008; Fernández and Wider, 2003).

1.4.2.3 Surface Plasmon Resonance

Surface plasmon resonance (SPR) is a method used to kinetically and thermodynamically quantify protein-protein, protein-nucleic acid and protein-small molecule interactions. The sensitivity levels of this technique have improved to the point that it can be used for fragment screening. SPR measures binding directly and requires the target protein to be immobilized on the surface of a gold chip. For FBS, a solution containing fragments is passed over the surface (Danielson, 2009; Neumann *et al.*, 2007). The physical basis for detecting binding is illustrated in **Figure 1.9**. The detection of binding in real time is based on SPR phenomena. Polarized light reflected off a gold surface is recorded by a detector in real time. Changes in the micro-environment of the gold surface will affect the reflection of light. Some SPR instruments are sensitive enough to detect binding of weakly interacting fragments (Dolezal *et al.*, 2013). An SPR-based experiment can provide a kinetic profile of binding and unbinding events (association k_a and dissociation rates k_d) (Cooper, 2002).
There are several methods for immobilizing protein on chip surface. Amine coupling is an example that is based on the formation of amide bonds between the protein and the chip surface. A potential shortcoming of this method is that the binding site of interest could be blocked during the immobilization process. An alternative strategy to amine coupling is to use the streptavidin-biotin interaction. This approach requires the covalent attachment of biotin at the N- or C-terminus of the protein. The biotinylated protein will bind to a streptavidin-coated chip surface with high affinity.

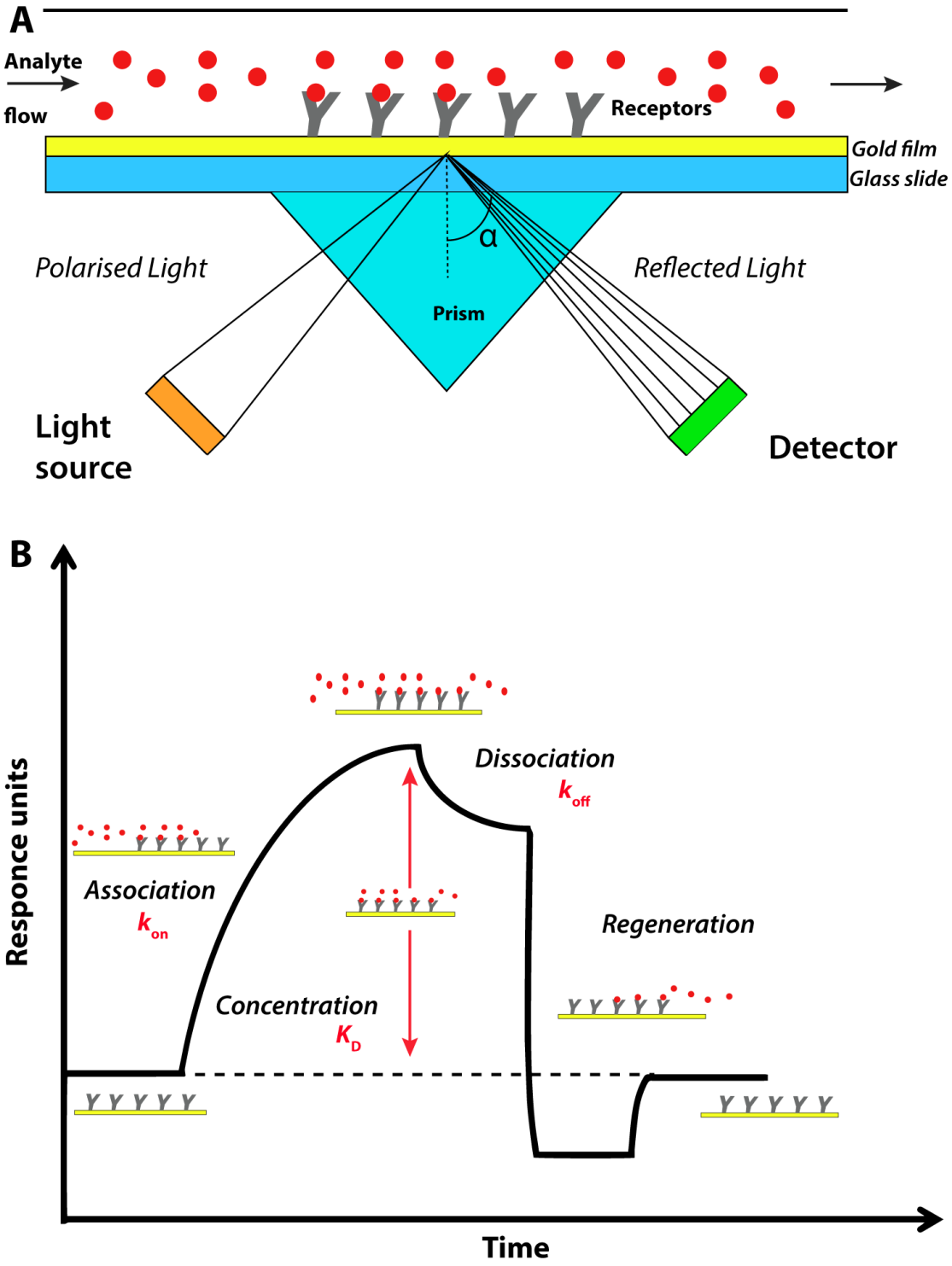


Figure 1.9. Schematic representation of a SPR experiment. (A) is the representation of the chip surface and (B) is the typical sensorgram corresponding to association/dissociation. Based on(Patching, 2014).

1.4.2.4 Isothermal Titration calorimetry

Isothermal titration calorimetry (ITC) is a valuable tool for measuring binding affinities, and thermodynamic data; the binding free energy (ΔG), which can be decomposed into

enthalpic and entropic contributions (ΔH and TΔS). When protein interacts with a ligand heat is either absorbed or released. In ITC, a ligand is injected into a chamber with the protein of interest, and the resulting change in heat is measured. ITC is able to measure interaction K_Ds ranging from mM to nM. This range can be extended to the pM range using competition assays. Heat changes can result from: ligand and macromolecule dilution, heat of mixing, buffer mismatch, in addition to the interaction of interest. The typical ITC instrument comprised of a reference cell and a sample cell. The temperature is in both chambers is constantly maintained. Heat changes in the sample chamber is measured relative to reference cell by the instrument (Holdgate and Ward, 2005).

A disadvantages of ITC is relatively low throughput and high sample consumption. It can, however, provide, binding stoichiometry as well as the aforementioned thermodynamic data on the basis of a single experiment (Doyle, 1997).

1.4.3 Fragment Optimisation

Fragment hits must be optimised in order to improve affinity. The common procedures for fragment optimisation include fragment merging, linking and/or growing (**Figure 1.19**).

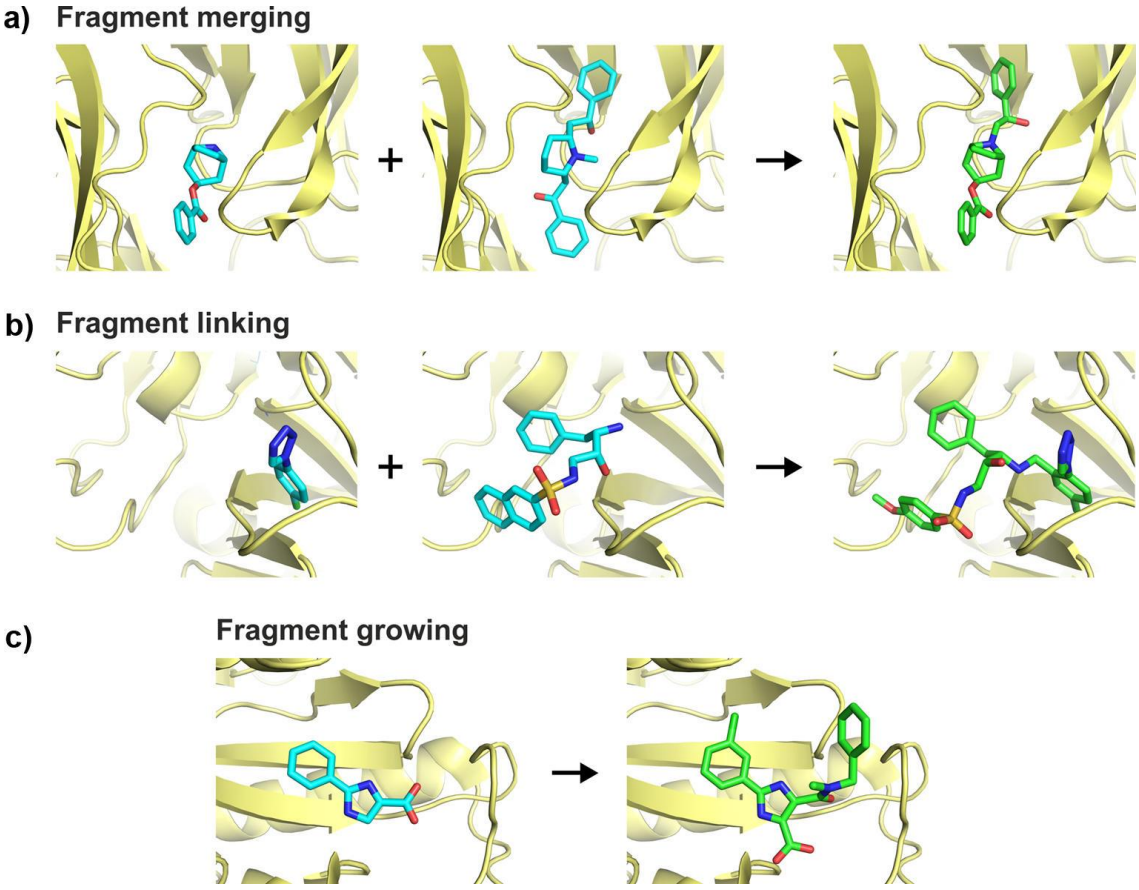


Figure 1.10. Illustration of fragment optimisation procedures: (a) fragment merging, (b) fragment linking and (c) fragment growing. Reproduced from (Scott et al., 2012).

The choice of optimisation path depends on the shape of the target binding pocket and the binding orientation of the fragments. In those instances where the binding pocket has a larger volume or consist of sub-pockets, a favourable approach may be to link fragments (assuming that there are fragments identified binding to multiple sub-pockets). Where pockets are small and narrow, the optimal choice is most likely to be fragment growth.

1.5 DNA Replication

Every dividing cell must pass on its genetic information by duplicating the genetic material, i.e. DNA. DNA replication is multi-step dynamic process entailing more than 30 different protein-protein and protein-nucleic acid interactions, the functions of which are coordinated by forming a large nucleoprotein complex that is known as the replisome (Schaeffer *et al.*, 2005).

In *E. coli*, DNA replication can be reduced to the following steps; (1) initiation, starting at *oriC*, the origin of replication, (2) elongation, in which two replication forks proceed in opposite directions around the chromosome and (3) termination, the step when replications ceases after replication forks reach a *ter* site.

1.5.1 The Origin of replication: Initiation

Initiation is the starting point of DNA replication, a process that is subdivided into three steps: (a) the recognition and unwinding of DNA at *oriC*, (b) the loading of the primosomal complex and RNA priming and (c) formation of DNA polymerase holoenzyme (Pol III HE). In order to initiate DNA replication, double-stranded DNA (dsDNA) must be unwound to form single-stranded DNA (ssDNA). Initiation begins with the interaction of DnaA protein with dsDNA at *oriC* to start the '*breaking*' of dsDNA. It creates two divergent bidirectional replication forks at *oriC* (Baker *et al.*, 1987).

The $DnaB_6$-$DnaC_6$ hexameric complex encircles ssDNA at *oriC* (Yuzhakov *et al.*, 1996; Bárcena *et al.*, 2001; Jezewska *et al.*, 1998). It provides ssDNA template for loading the DnaG primase to form $DnaB_6$-$DnaG_3$ primosomal complex (Fang *et al.*, 1999).

The final step of initiation includes the formation of Pol III HE assembly. Poll III HE is a large complex consisting of the clamp loader complex (τ, γ, δ, δ', χ and ψ subunits), Pol III catalytic core (α, ε and θ subunits) and sliding clamp (β_2) for coordination of strand synthesis (McHenry and Crow, 1979; McHenry, 2011; O'Donnell and Studwell, 1990; Stukenberg and O'Donnell, 1995). A schematic view of initiation is shown in **Figure 1.11A**.

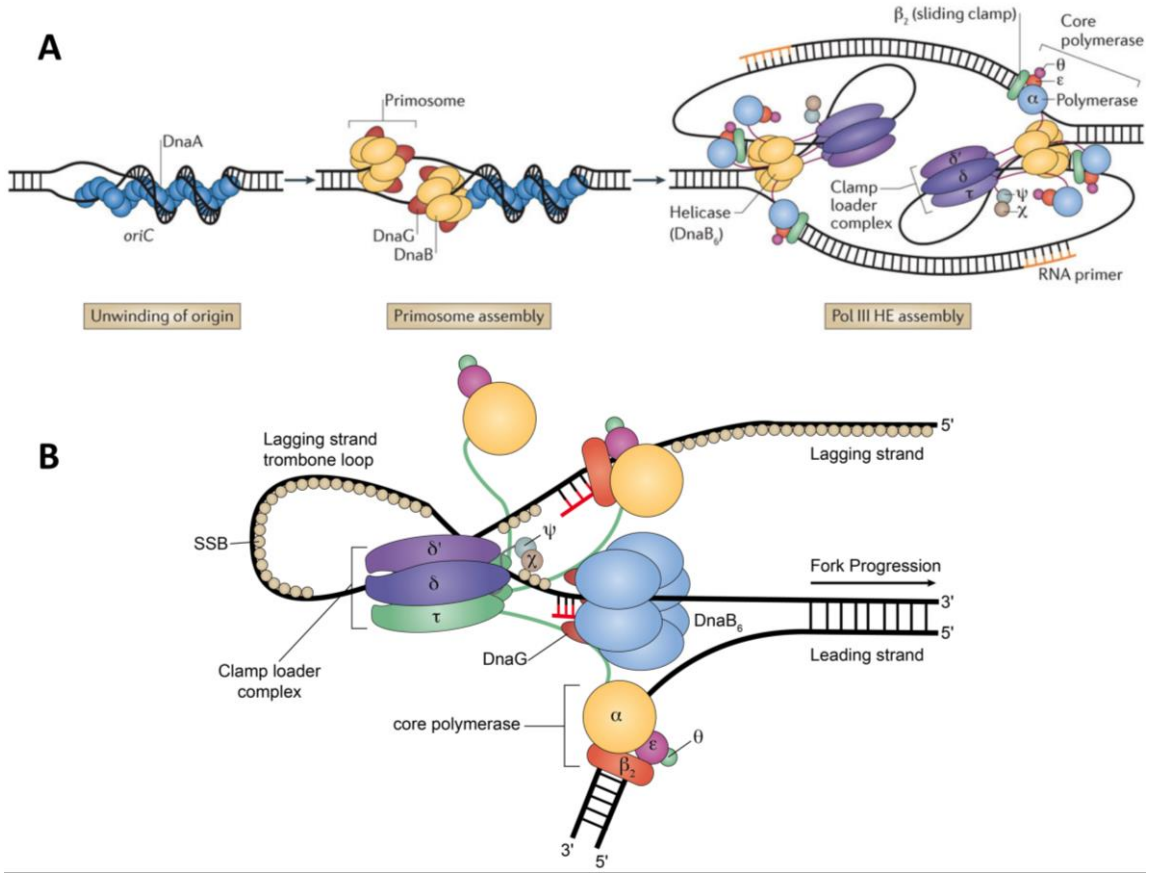


Figure 1.11. A schematic representation of DNA replication. (A) Steps of initiation and (B) replication fork. Modified from (Robinson and van Oijen, 2013).

1.5.2 Elongation

Once all necessary components are assembled at *oriC*, elongation (the synthesis of new DNA strands) starts. Elongation is a highly synchronized process that involves the unwinding of dsDNA and the continuous synthesis of the leading and lagging strands.

Due to the antiparallel nature of parental DNA, one strand is synthesized continuously (the leading strand), the other discontinuously (the lagging strand). The lagging strand is comprised of so-called "Okazaki fragments" (Okazaki *et al.*, 1967). For the leading strand synthesis, DnaG primase synthesise the RNA primers, thereafter to be used by polymerase to synthesise a new strand from the leading strand template (Kohara *et al.*, 1985). Meanwhile for lagging strand, multiple primers are required for discontinuous synthesis of Okazaki fragments (McMacken *et al.*, 1977). Primer synthesis requires contact between the DnaG primase and DnaB helicase (Arai and Kornberg, 1979). The complete schematic representation of replisome at the replication fork is shown in **Figure 1.11B**. For leading strand synthesis to take place, a few priming steps may be sufficient for continuous strand synthesis to occur. In order for a complete synthesis of

the lagging strand, efficient priming that guarantees the completion of newly synthesized strand is required. At each step DnaG primase is recycled. Subsequent to synthesized Okazaki fragment synthesis, Pol I is required to remove the RNA primer and replace it with DNA, and DNA ligase to join the fragments together. The mechanism by which primase is recycled has been explained using a three-point switch model (Yuzhakov *et al.*, 1999). It involves DnaB helicase, DnaG primase, single stranded DNA-binding protein (SSB), and the χ subunit of Pol III. The SSB coating ssDNA interacts through its C-terminal region with DnaG primase and Pol III χ subunit (Shereda *et al.*, 2008). DnaG primase, in turn, is bound to DnaB helicase. DnaG primase requires SSB-Ct binding to maintain contact with the primed site. Pol III χ displaces DnaG primase, which can then be recycled. The recycling of DnaG primase is repeated for each Okazaki fragment.

1.5.3 Termination

The two replication forks eventually meet at the one of the specific replication terminus (*Ter*) sites in the chromosome. The Tus protein binds to *Ter* sites and block replication forks that approach from non-permissive sides causing termination of replication (Coskun-Ari *et al.*, 1994; Coskun-Ari and Hill, 1997).

1.5.4 Single Stranded DNA-Binding Protein

SSB is one of the essential components of DNA replication, replication restart, recombination and DNA repair. It is comprised of oligonucleotide/oligosaccharide binding (OB) domain (residues 1-114), responsible for binding to the ssDNA and a flexible C-terminal region (residues 115-177) that terminates with a highly conserved hexapeptide motif (SSB-Ct; residues 168-177) at the end of the C-terminus (**Figure 1.12**) (Shereda *et al.*, 2008).

Figure 1.12. Schematic view of E. coli SSB. Dotted lines represent the flexible region. Increasing font size represents greater degree of conservation of residues among SSB homologs from different species.

SSB is a tetramer in solution. The main function of SSB is to protect ssDNA from degradation. NMR and X-ray crystal structures demonstrate that in full-length SSB, the

C-terminal domain including the highly conserved C-terminal motif is disordered (Shishmarev *et al.*, 2014).

Two DNA-binding modes are identified for ssDNA to SSB: the 35-mode (35 nucleotide based are wrapped around SSB tetramer) and 65-mode (65 nucleotide residues are bound to SSB). Each binding mode depends on experimental conditions such as salt concentration, temperature, pH and protein concentration (Raghunathan *et al.*, 1997). The crystal structure of *E. coli* SSB alone and in complex with ssDNA in the $(SSB)_{35}$ binding mode are is shown in **Figure 1.13**.

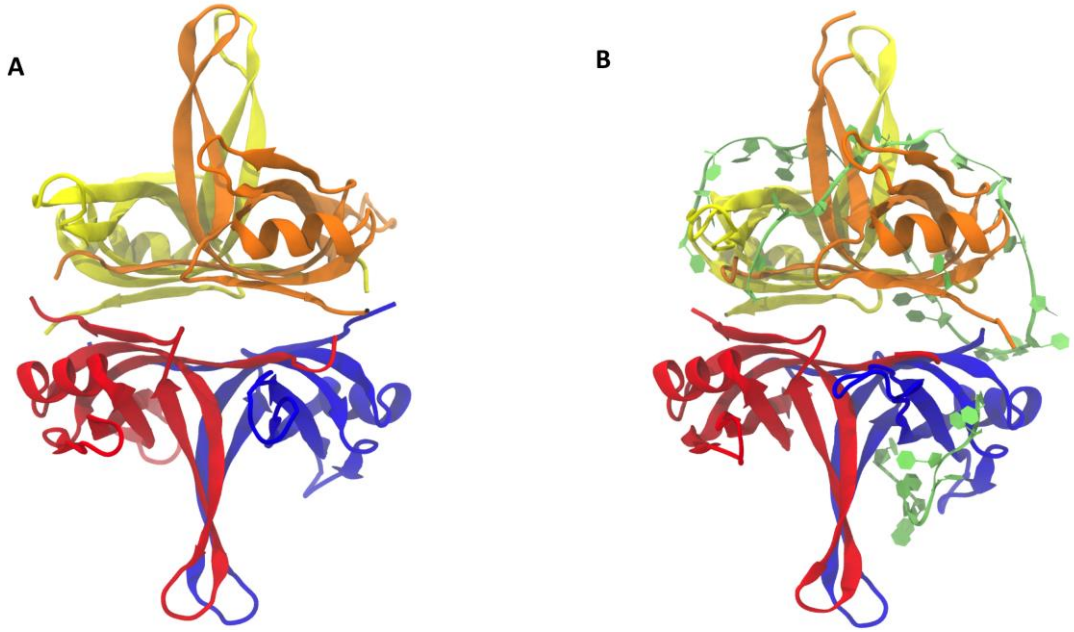


Figure 1.13. Crystal structure of *E. coli* SSB (A) and with ssDNA in 35 mode (B). Each monomer represented in different colour. ssDNA is shown in green (PDB ID: 4MZ9 and 1EYG).

The availability of structures of SSB homologues from different species (*E. coli, P. falciparum, M, tuberculosis*) allows conserved structural and functional features of SSB/ssDNA interactions to be identified. Comparison of *E. coli* and *Deinococcus radiodurans* structures supports the overall ssDNA binding mechanism (George *et al.*, 2012). *Plasmodium falciparum* is a parasitic protozoan responsible for malaria. *P. falciparum* SSB is also a tetramer with different amino acid composition at the C-terminus. The structures of *Pf*-SSB tetramer and its complex with a ssDNA shows similarities to *E. coli* SSB, however, ssDNA wraps around SSB with opposite handedness to that observed in *E. coli* (**Figure 1.14**) (Antony *et al.*, 2012).

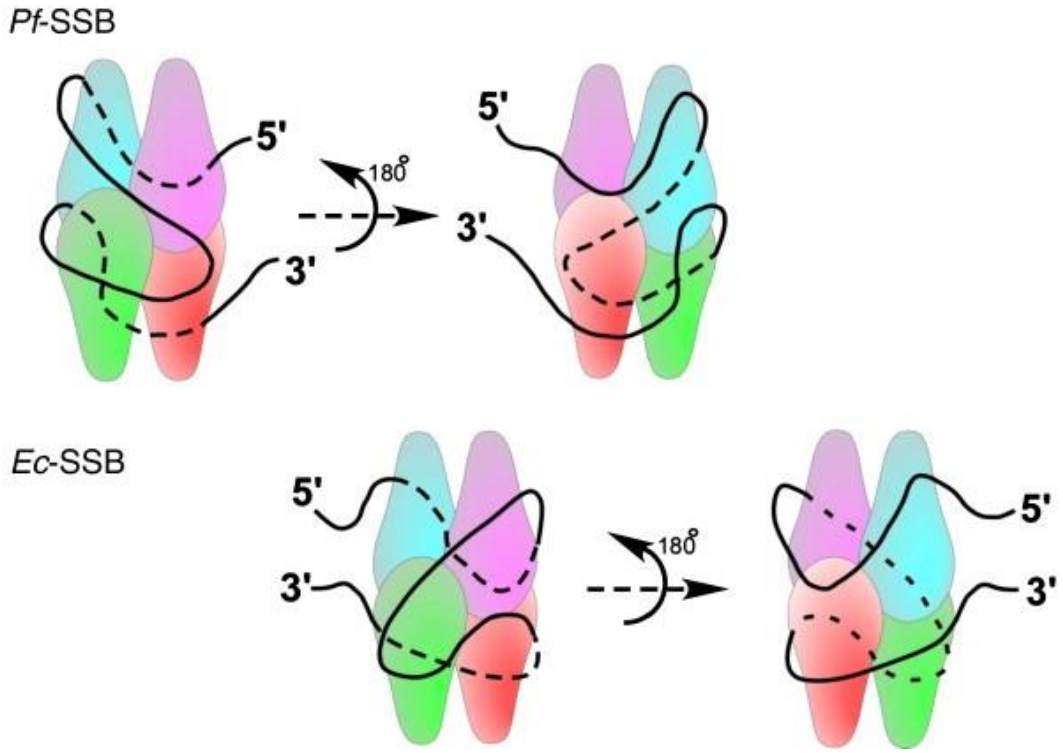


Figure 1.14. Different binding modes of SSB and ssDNA in E. coli and Pf. respectively. Reproduced from (Antony et al., 2012).

SSB is a protein-interaction hub (Shereda *et al.*, 2008) that interacts with at least 14 proteins. It was shown in some cases that these the interactions occurs *via* the conserved C-terminal hexapeptide motif. A list of SSB-Ct binding partners is shown in **Table 1.3**.

Table 1.3. SSB-Ct binding partners. Structures of proteins where complexes with SSB-Ct have been determined are indicated (*).

Protein	**SSB-Ct required?**	**Binding affinity K_D**	**Function**	**References**
χ subunit of DNA Pol III)*	Yes	9 μM	Part of clamp loader complex	(Kelman *et al.*, 1998; Yuzhakov *et al.*, 1999; Witte *et al.*, 2003; Butland *et al.*, 2005)
DnaG*	Yes	2 μM	RNA primers synthesis	(Yuzhakov *et al.*, 1999)
ExoI*	Yes	136 nM	Methyl directed mismatch repair	(Molineux and Gefter, 1974; Butland *et al.*, 2005; Lu and Keck, 2008)
ExoIX	Unknown	-	Phospodiester backbone cleavage	(Hodskinson *et al.*, 2007)
PriA*	Yes	2 μM	Replication restart	(Cadman and McGlynn, 2004; Butland *et al.*, 2005; Lecointe *et al.*, 2007)
PriB	Unknown	-	Replication restart	(Low *et al.*, 1982)
RecG	Yes	n.d.	Replication and repair	(Butland *et al.*, 2005; Lecointe *et al.*, 2007; Buss *et al.*, 2008)
RecJ	Unknown	-	Replication and repair	(Butland *et al.*, 2005; Han *et al.*, 2006)
RecO*	Yes	60 nM	Recombination	(Umezu and Kolodner, 1994; Ryzhikov *et al.*, 2011)
RecQ	Yes	6 μM	Replication and repair	(Butland *et al.*, 2005; Lecointe *et al.*, 2007; Shereda *et al.*, 2007; Shereda *et al.*, 2009)
Topoisomerase III	Unknown	-	Replication and repair	(Butland *et al.*, 2005)
DNA Pol II	Yes		Replication and repair	(Molineux and Gefter, 1974)
DNA Pol V	Yes		Replication and repair	(Arad *et al.*, 2008)
UDG*	Yes	1.7 nM	Replication and repair	(Purnapatre *et al.*, 1999; Handa *et al.*, 2001)
DNA Pol IV	Yes		DNA strain relaxation during the unwinding	(Furukohri *et al.*, 2012)

The numerous essential interactions SSB-Ct makes with other proteins, makes is a potential drug target. Furthermore, the human SSB C-terminus is distinctive and therefore inhibitors of bacterial SSB-Ct interactions are unlikely to react with human SSB-Ct-binding partners (Robinson *et al.*, 2010).

1.5.5 DnaG primase

DnaG primase is a DNA-dependent RNA polymerase encoded by the *dnaG* gene (Bouché *et al.*, 1975). The *E. coli* protein consist of 581 amino acids (65.5 kDa) with three distinct domains (**Figure 1.15**): an N-terminal zinc-binding domain (ZBD, residues 1-110); the central domain RNA polymerase domain (RPD, residues 111-433); and the C-terminal helicase-binding domain (HBD, DnaGC, residues 434-581).

Figure 1.15. Schematic view of *E. coli* DnaG primase. Each domain is shown in a different colour.

The structure of all three domains in isolation have been determined. The structure of the ZBD of DnaG primase from *B. stearothermophilus* was solved by X-ray crystallography (Pan and Wigley, 2000). The zinc-binding motif contains conserved Cys-His-Cys-Cys motif responsible for coordinating a Zn^{2+} ion (Berg, 1990). The ZBD is responsible for binding to DNA templates, specifically short nucleotide sequences [5'-C(A/T)G-3']. Once the contact is established with ssDNA, the RPD domain binds and synthesizes RNA primers (Kusakabe and Richardson, 1996; Anon, 2000) (**Figure 1.16**).

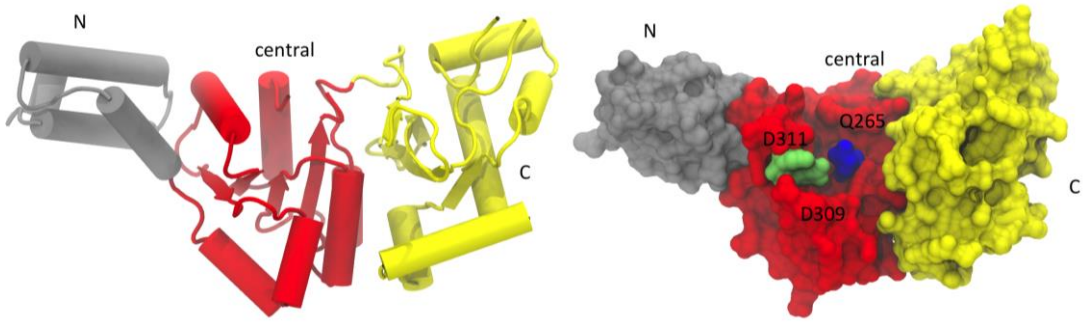


Figure 1.16. Crystal structure of *E. coli* RPD domain of DnaG primase (PDB ID:1DD9) (Keck et al., 2000). Grey, red and yellow colour of the DnaG-RNAP domain correspond to N-terminal (residues 115 to 240), central (residues 241 to 367), and C-terminal (residues 368 to 428) segments. Residue Q265 is coloured blue and the Asp-X-Asp residues are coloured green.

The catalytic core of DnaG has a conserved stretch of ~80 residues (within the central domain; residues 259-341) identified in type IA and II topoisomerases and in DnaG-type primases in archaea and bacteria, called a TOPRIM (topoisomerase/primase) sub-domain found in the central segment of RPD (**Figure 1.16**). There are only five conserved residues in all of these three types of proteins. In *E. coli*, two of them are glycine residues and the other three are acidic residues (Glu265, Asp309 and Asp311). Both types of topoisomerases have a conserved tyrosine residue in their TOPRIM domains responsible for binding DNA that is missing in the primase. Primase contains a tyrosine residue (Tyr197) with an equivalent function that is not contained in its TOPRIM domain lying in a position equivalent to the aforementioned tyrosine residue in the topoisomerases to bind to the DNA. TOPRIM domains have two conserved acidic motifs (Glu265 and residues 309-311 in *E. coli* DnaG) (**Figure 1.16**). It has been shown that an essential conserved motif is required for Mg^{2+} mediated NTP-binding (Aravind *et al.*, 1998; Keck and Berger, 2000; Podobnik *et al.*, 2000).

DnaG primase interacts with DnaB helicase (Bailey *et al.*, 2007; Loscha *et al.*, 2004) and SSB (Yuzhakov *et al.*, 1999; Naue *et al.*, 2013) through its C-terminal domain (DnaGC). DnaGC is a monomer in solution (Oakley *et al.*, 2005). The structure of *E. coli* DnaGC was previously determined by X-ray crystallography at 2.8 Å resolution and NMR spectroscopy (**Figure 1.17**) (Loscha *et al.*, 2004).

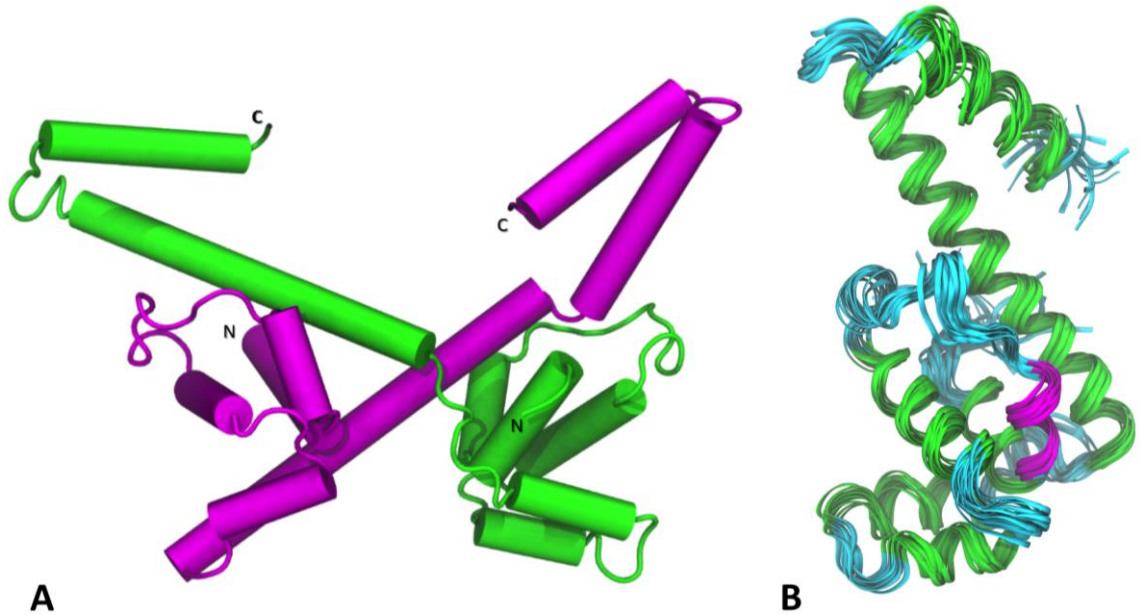


Figure 1.17. (A) Crystal structure of DnaGC (PDB ID: 1T3W) (Oakley et al., 2005). Ribbon representation of two molecules in the asymmetric unit. Each monomer is shown in a different colour (green and magenta). (B) NMR structure of DnaGC (PDB ID: 2HAJ) (Su et al., 2006). The 20 lowest energy conformations are shown.

The crystal structure contained two monomers in the asymmetric unit that form a domain-swapped dimer (**Figure 1.17A**). The first 14 residues of the construct (residues 429 to 443) were not observed in the crystal structure and were furthermore observed to be flexible by the NMR (Loscha *et al.*, 2004; Su *et al.*, 2006). A recent crystal structure of the complex between the hexameric DnaB and the DnaG primase of *G. stearothermophilus* was solved by X-ray crystallography (**Figure 1.18**) (Bailey *et al.*, 2007). The crystal structure showed that DnaB interacts with DnaGC forming stable two-layered ring structure, interacting through DnaGCs C-terminal helix.

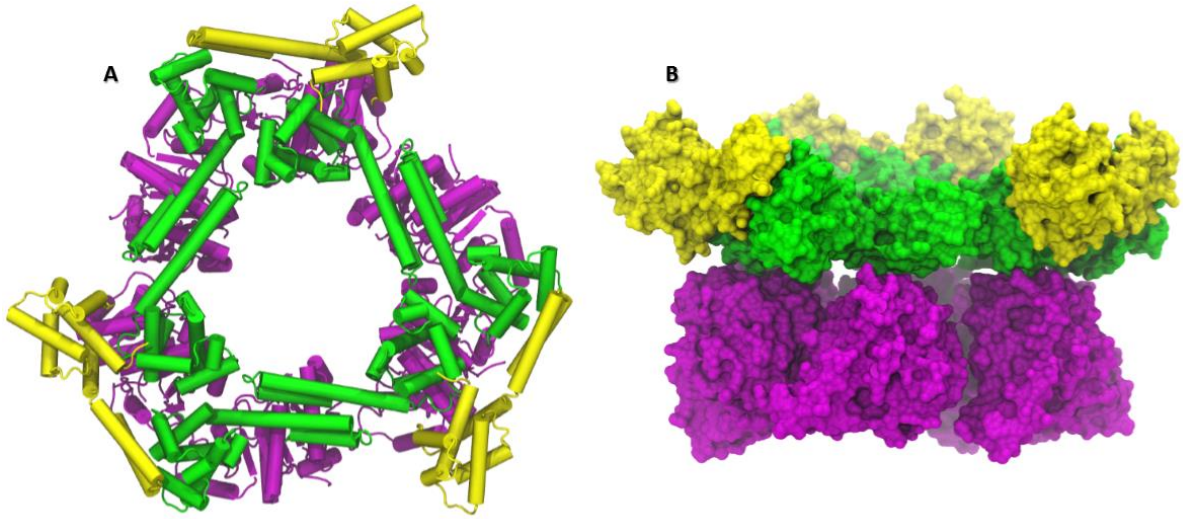


Figure 1.18. Structure of the complex of hexameric DnaB helicase and the HBD of DnaG from *G. stearothermophilus* (PDB ID: 2R6A) (Bailey et al., 2007). (A) Cartoon representation of the structure from the top with three monomers of DnaGC (yellow) biding to the N-terminal domain (green) C-terminal domain (magenta) of hexameric DnaB helicase, (B) Surface representation of the structure from the side with same colour representation as in (A).

The stoichiometry and binding affinity of the DnaB-DnaG interaction was measured by SPR (Oakley *et al.*, 2005), and confirmed by the crystal structure of the primosome complex from *G. stearothermophilus*. The interaction of three DnaGC molecues with hexameric DnaB helicase may increase the processivity of replication (Bailey *et al.*, 2007). Modelling suggests that *E. coli* DnaG/DnaB interactions are similar (Loscha *et al.*, 2004; Tak Lo, 2012).

Recently, the structure of the complex of DnaGC with the SSB-Ct peptide was determined (Tak Lo, 2012). The crystal structure of chimeric DnaGC-linker-SSB-Ct construct was solved at 1.5 Å resolution, providing insight into the key interactions between SSB-Ct and DnaGC. The SSB-Ct peptide interacts with the helical bundle sub-domain of DnaGC. A binding pocket comprised of Leu446, Met451, Ile455, Pro480, Leu484, Thr515 and Leu519 residues, forming a hydrophobic cavity to accommodate

the last 3 residues of SSB-Ct peptide (Ile175, Pro176 and Phe177) (Tak Lo, 2012) (**Figure 1.19**).

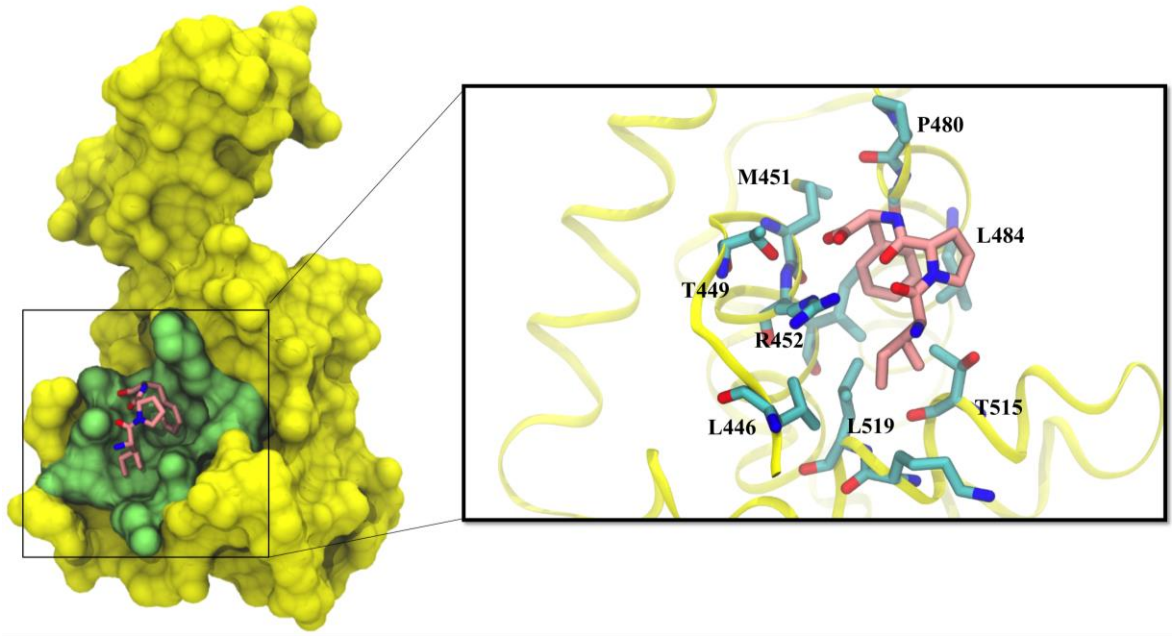


Figure 1.19. Crystal structure of DnaGC-SSB-Ct chimera. DnaGC is shown with a yellow surface with the SSB-Ct binding pocket in green. SSB-Ct peptide is represented in sticks. Key residues forming the pocket is shown in right (Tak Lo, 2012).

Comparison of the SSB-Ct-binding pocket on DnaGC with that of other SSB-binding proteins highlights some similarities among the SSB-Ct binding partners. The last residue of the SSB-Ct (Phe177) is buried in the pocket, forming hydrophobic contact The carboxylic acid group is involved in electrostatic and hydrogen bond interactions with the DnaGC pocket, most of these with the Arg452 guanidinium group. Mutagenesis studies carried out simultaneously by two different groups confirm that some of the residues in the pocket contribute to binding, as demonstrated by K447A, T450A, R452A and K518A mutants that show severely attenuated binding (Naue *et al.*, 2013; Tak Lo, 2012). Work published recently has confirmed the location of the binding pocket of SSB-Ct peptide on DnaGC using NMR (Naue *et al.*, 2013). The perturbed residues of DnaGC in ^{15}N-HSQC spectra upon titration of SSB-Ct peptide were mapped onto protein surface, identifying the SSB-Ct binding pocket (**Figure 1.20**).

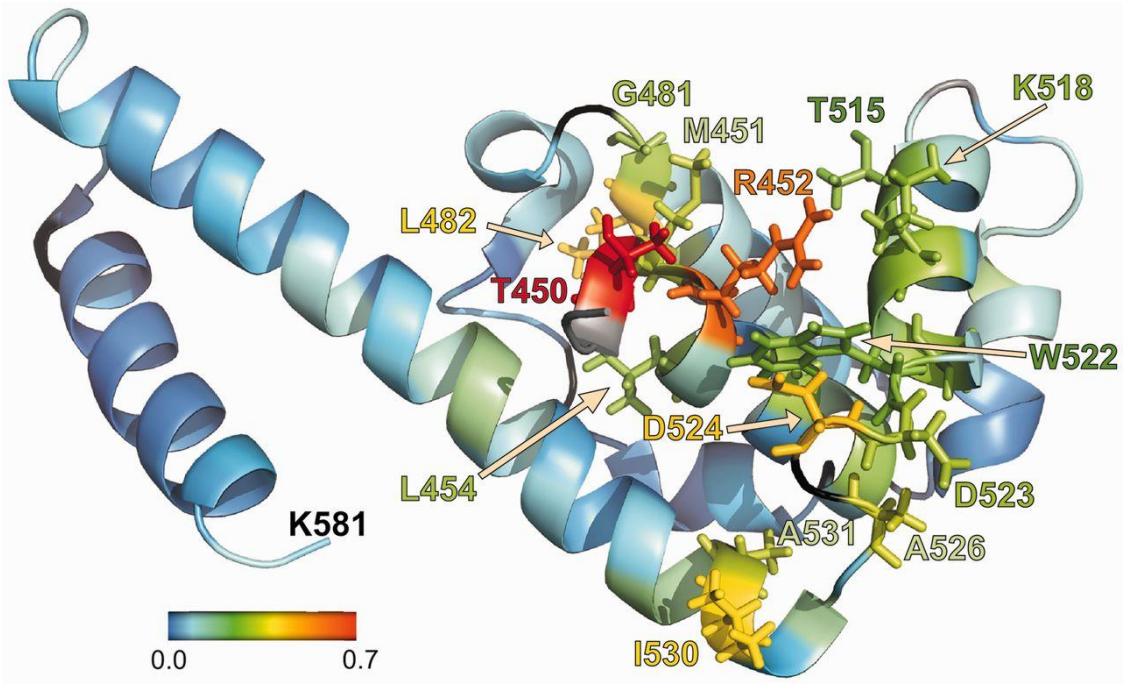


Figure 1.20. Binding surface identified by NMR titration of SSB-Ct peptide into ^{15}N-labelled DnaGC. Protein is coloured according to their chemical shift magnitude upon addition of peptide. Blue represents no chemical shifts, red is highest chemical shift observed. Residues forming the SSB-Ct binding pocket are in stick representation. Reproduced from (Naue et al., 2013).

1.6 Scope of this work

The inhibition of SSB-Ct-binding activity could be used to treat pathogenic bacteria resistant to currently available antibacterials (Robinson *et al.*, 2010). This thesis considers the interactions of SSB and its binding partner interactions. MD simulations were conducted in order to understand the dynamic behaviour of SSB-Ct in complex with its binding partners using currently available crystal structures of SSB-Ct complexes (DnaGC, pol III χ, RecO and ExoI) (**Chapter 3**).

Fragment-based screening against DnaG primase was carried out (**Chapter 4**). The first generation of hits identified by SPR and STD experiments were further characterised by protein-based NMR approaches (**Chapter 5**). *In silico* hit-to-lead optimisation was carried out (**Chapter 6**). Finally, the first generation of optimised hits were tested against other SSB-Ct binding partners.

Chapter 2

Material and Methods

2.1 MATERIALS

2.1.1 E. coli Strains and Plasmid Vectors

Plasmids (PKL1176) encoding *E. coli* DnaGC and RCD proteins (**Figure 2.1**) were described previously (Loscha *et al.*, 2004; Tak Lo, 2012). The *E. coli* BL21(λDE3)*recA* strain was used as an expression host for the overproduction of DnaGC and RCD proteins (Loscha *et al.*, 2004).

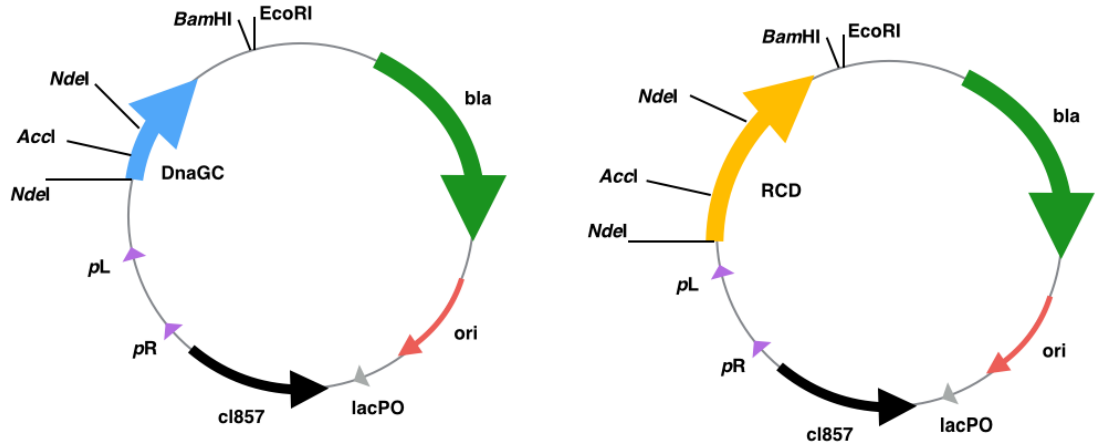


Figure 2.1. Schematic diagram of DnaGC and RCD plasmids (Loscha et al., 2004; Tak Lo, 2012).

2.1.2 Growth Media

2.1.2.1 LB Liquid Medium

Lysogeny broth (LB) media (25 g/L) (Moore and Brubaker, 1975) supplemented with 25 mg/L thymine and 100 mg/L Ampicillin was used for growth of *E. coli* strains. Media was autoclaved prior to addition of antibiotic before inoculation with bacteria. Overnight cultures (5 mL) were used to inoculate fresh LB cultures (1 L) for overproduction of protein. *E. coli* strains with λ vectors were grown at 30 °C with shaking at ~260 rpm using an orbital shaker. Overnight growth (~18 h inoculation) was used.

2.1.2.2 LB Solid Medium

Cells containing targeting vectors were grown on LB agar plates (15 g/L agar in LB media with appropriate antibiotic) in an incubator at 30 °C overnight (~ 18 h).

2.1.3 Custom Peptides

Peptides for fluorescence polarization assays were obtained from Mimotopes (Clayton, VIC). Peptides were labelled with fluorescein at their N-termini. Prior to use, peptides were dissolved in 100% DMSO and stored at -20 °C.

N-terminally biotinylated SSB-Ct peptide [Biotin-(Ahx)-GS**APSNEPPMDFDDDIPF**; where Ahx an amino-hexanoate spacer, is followed by 16 C-terminal residues of SSB highlighted in bold] was synthesized at Biomolecular Resource Facility at the John Curtin School of Medical Research, ANU. The peptide was dissolved in water and stored at -20 °C.

2.1.4 Protein purification components

2.1.4.1 ^{15}N-labelledMinimal Media

Minimal media contained ^{15}N-labelled ammonium chloride, thymine, glucose, solution of trace metals and basic salt solution. The full recipe for minimal media is shown in **Table 2.1**.

Table 2.1. Composition of minimal media.

20 × Basic salt solution	**Per 0.5 L**
$NaH_2PO_4 \times 2H_2O$ (pH 7.0)	61 g
K_2HPO_4	106 g
1000 × Trace metal solution	**Per 100 mL**
MQW	36 mL
0.1M $FeCl_3$ in 0.1 M HCl	50 mL
1 M $CaCl_2$	2 mL
0.1 M Na_2SeO_3	2 mL
0.1 M H_3BO_4	2 mL
1 M $MnCl_2$	1 mL
1 M $ZnSO_4$	1 mL
0.2 M $CoCl_2$	1 mL
0.1 M $CuCl_2$	2 mL
0.2 M $NiCl_2$	1 mL
0.1 M $NaMoO_2$	2 mL

The mixture of 50 mL of 20 × basic salt solution with 1 g/L $^{15}NH_4Cl$ in 1L water was autoclaved. Prior to inoculation of 1L culture, 1 mL of 1 M $MgSO_4$ with 1 mL of metal mix and 3 g of D-(+) glucose with 100 µL of 10 g/L thymine was added before transferring to 30 °C in presence of 100 µg/mL Ampicillin.

2.1.4.2 Buffers for Protein Purification

Buffers used for purification of RCD and DnaGC were: lysis buffer [50 mM Tris-HCl (pH 7.6), 1 mM EDTA, 1 mM DTT, 10% (w/v) sucrose, 150 mM NaCl, 10 mM spermidine]; Buffer A [50 mM Tris-HCl (pH 7.6), 1 mM EDTA, 1 mM DTT]; and Buffer B [50 mM Tris-HCl (pH 7.6), 1 mM EDTA, 2 M NaCl and 1 mM DTT].

2.1.5 Fragment library

The "first pass screen" fragment library (Zenobia Therapeutics) was used for the SPR competition assay. For storage, stocks for each fragment were made by addition of 50 µL DMSO to each fragment (200 mM final concentration). Each fragment was diluted with DMSO to 1 mM final concentration for the competition assay. The compounds were assayed for chip surface binding in order to eliminate false positives.

The MIPS fragment library was used for STD NMR experiments. It comprised of around 1140 fragments. The individual fragments were diluted in $^2H_6$-DMSO to give ~660 mM final stock concentration (Doak *et al.*, 2013).

The fragments were mixed in cocktails of up to 6 compounds and had well-resolved resonances in their 1D ^1H NMR spectra. Fragment cocktails were generated such that overlaps in their 1D ^1H NMR spectra were avoided.

2.2 METHODS

2.2.1 General Protein Biochemistry Methods

2.2.1.1 Dialysis of Protein

Dialysis was used for protein buffer exchange during the purification steps and prior to other experiments. Spectra/Por standard regenerated cellulose dialysis membranes (spectrum Laboratories, Rancho Dominguez, CA, USA) were used for protein dialysis. An appropriate molecular weight cut-off was selected for each protein to be smaller

than molecular weight of the target protein. Three changes of new buffer were performed each one at least 3 h duration at 4 °C.

2.2.1.2 Fast Protein Liquid Chromatography Purification of Protein

Overexpressed proteins were purified by DEAE, MonoQ ion exchange columns using the ÄKTA™ system (ÄKTApurifier™ or ÄKTAFPLC™) instrument equipped with fraction collector (Frac-9500 or Frac-920), controlled by UNICORN™ software (GE Healthcare). Protein elution was assayed continuously by protein absorbance at $\lambda = 280$ nm and conductivity. All protein purification steps were carried out at 4 °C.

2.2.1.3 Denaturing Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis

Protein purity was assessed using sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE). Pre-cast gels (4-12%) (Bio-Rad, NSW Australia) were used in all experiments. Protein molecular weight markers were obtained from Bio-Rad (Precision Plus Protein^{TM} Dual Colour standards). Protein samples were mixed with equal volume of the dye [300 mM Tris base, 15% (v/v) glycerol, 0.6% (w/v) bromophenol blue, 200 mM DTT, 1% SDS]. Samples were loaded onto gels, and 180-200 V was applied until the bromophenol blue dye had reached the end of the gel. Gels were then washed with water and stained by adding staining solution [0.2% (w/v) Coomassie brilliant blue R, 40% (v/v) methanol and 10% (v/v) acetic acid] with heating in a microwave over for 1 minute (until boiling) and gently shaken for ~15 min. Gels were destained [40% (v/v) isopropanol, 10% (v/v) acetic acid] until the background colour was gone.

2.2.1.4 Determination of Protein Concentration

Protein UV absorbance at 280 nm wavelength (A280) was used to determine protein concentration in solution using NanoDrop 2000c™ (Thermo Scientific). The proteins extinction coefficient at 280 nm (ε_{280}) was calculated by *ExPASy Server* using the number of tryptophan and tyrosine amino acids in the protein sequence (Gasteiger *et al.*, 2005).

2.2.1.5 Concentration of Protein: Ultrafiltration

Protein solutions were concentrated using Amicon Ultra-4 centrifuge filters with 3 kDa molecular weight cut-off (Millipore, Bedford, USA). Prior to use, membranes of the filters were rinsed twice with Milli-Q water to remove preservative (glycerol). Solutions were centrifuged for 15 min at 4000 × g in a Sorvall® Super T21 Benchtop Super speed Centrifuge (Thermo Scientific, Waltham, MA) using a ST-H750 rotor. Following each centrifugation step, samples were thoroughly mixed by repeated pipetting to prevent protein aggregation.

2.2.1.6 Electrospray Ionization Mass Spectrometry (ESI-MS)

The molecular weight of purified protein was confirmed using ESI-MS that gives the molecular mass of protein to 0.01% precision. Mass spectra were acquired using quadrupole-time-of-flight (Q-TOF) Ultima™ mass spectrometer equipped with Z-spray probe and mass analyzer with 32,000 *m/z* range. The instrument was controlled using MassLynx 4.1 software (Micromass, UK), and calibrated using cesium iodide (10 mg/mL in 70% isopropanol). The protein concentration required for measurements was ~10 μM. Proteins were dialyzed against 0.1% of formic acid prior to analysis.

2.2.1.7 Expression and Purification of the RCD and DnaGC domains

The RCD overproduction and purification protocol was essentially the same as unlebelled DnaGC described previously (Loscha *et al.*, 2004). Plasmid pKL1176 containing gene encoding the RCD domain was transformed into *E. coli* BL21(λDE3)*recA* cells (Williams, 2002). Cells were grown in 1 L cultures in LB broth in the presence of 100 μg/mL Ampicillin at 30 °C until OD_{600} reached ~0.6. The temperature was rapidly shifted to 42 °C by transfer of flasks to a heated incubator, and the cultures shaken for another 3 h. Cells were harvested by centrifugation (11,000 × g, 15 min), the cell pellet was resuspended in lysis buffer (15 mL lysis buffer per 1 g of cells). Cells were lysed by passing the cell suspension through a French press. After centrifugation (10,000 × g, 90 min), 0.3 g/mL $(NH_4)_2SO_4$ was used to precipitate the proteins from the supernatant, followed by dialysis against 2 L of buffer A with 100 mM NaCl, changed twice. The protein was passed twice though a DEAE column (DEAE-650M resin), first in the presence of buffer A with 100 mM NaCl, and second in buffer A without NaCl. A third anion exchange step utilized a Mono Q™ column

(10/100 GL column 8 mL: 10 × 100 mm: GE Healthcare) with the same buffer conditions (buffer A without NaCl) followed by gel filtration (Superdex 75; GE Healthcare) using the buffer A with 100 mM NaCl. Samples were clarified by centrifugation (40,000 × g, 15 min) prior loading onto columns.

The overproduction of ^{15}N-labelled DnaGC in ^{15}N-rich minimal media was as described for unlabelled DnaGC/RCD protein.

Cysteine mutants of DnaGC were made by Dr. Zhi-Qiang Xu. Cysteine mutants of ^{15}N-labelled DnaGC were expressed and purified as described above for ^{15}N-labelled DnaGC.

2.2.2 Computational Chemistry Methods

2.2.2.1 System Preparation

Starting coordinates for MD simulations were derived from crystal structures of DnaGC/SSB-Ct complex at 1.5 Å resolution (Tak Lo, 2012), the RecO/SSB-Ct complex at 2.3 Å resolution (PDB ID: 3Q8D) (Ryzhikov *et al.*, 2011), the ExoI/SSB-Ct complex at 2.7 Å resolution (PDB ID: 3C94) (Lu and Keck, 2008) and the χ/ψ-SSB-Ct complex at 1.85 Å resolution, where the SSB-Ct peptide was modified to WDIPF (PDB ID: 3SXU) (Marceau *et al.*, 2011).

Where multiple molecules were found in the asymmetric unit, only the first was used. For all available structures, the SSB-Ct peptide was extended where necessary by manually model building in COOT (Emsley and Cowtan, 2004) to include the last 9 residues of that protein, i.e. MDFDDDIPF. All water molecules were removed. Hydrogen atoms were built using the PSFGEN plugin in VMD (Humphrey *et al.*, 1996). Histidine protonation states were assigned based on visual inspection of their hydrogen-bonding environments. In all systems the CHARMM27 protein force field was used (MacKerell *et al.*, 1998; Mackerell *et al.*, 2004). Water molecules (TIP3P) were placed around protein models so as to create rectangular boxes (Jorgensen *et al.*, 1983). In all cases there was a minimum distance of 7 Å between protein and the edge of the box. Sodium and chloride ions were added to ensure that the systems had no net charge. Four systems: DnaGC/SSB-Ct (~7,700 atoms), RecO/SSB-Ct (~22,500 atoms), ExoI/SSB-Ct (~42,500 atoms) and χ/ψ–SSB-Ct (~28,200 atoms) were thus generated (**Figure 2.2**).

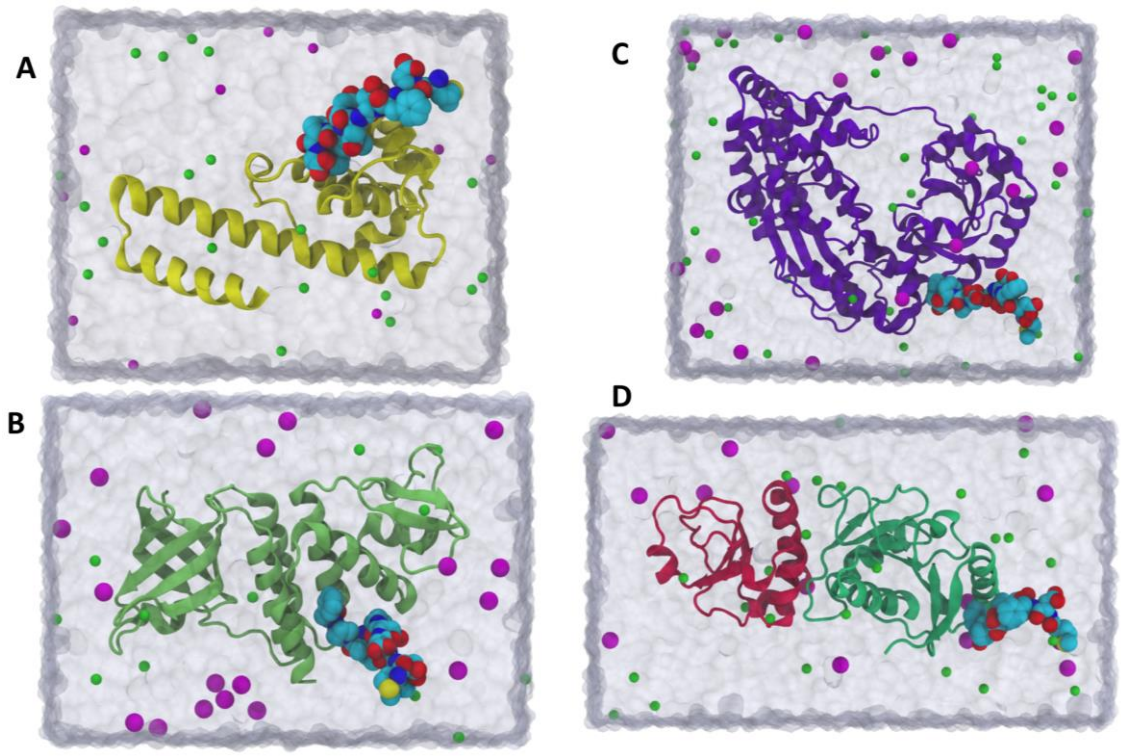


Figure 2.2. Systems od SSB-binding partners: DnaGC (A), RecO (B), ExoI (C) and χ/ψ (D) in water box in presence of sodium (magenta) and chloride (green) ions.

2.2.2.2 Molecular Dynamics Simulations Protocol

Before starting the MD simulations, the systems were subject to energy minimization for 10,000 steps. Constant temperature (310 K) was maintained using Langevin dynamics (Izaguirre *et al.*, 2001; Bussi and Parrinello, 2007) with a 5 ps^{-1} damping constant applied to non-hydrogen atoms. Periodic boundary conditions were used with the Nosé-Hoover Langevin piston pressure control method (Feller *et al.*, 1995) with 100 fs piston period and 50 fs decay rate, respectively, to maintain a constant pressure of 1.013 Bar. The Particle-mesh Ewald (PME) (grid resolution < 1 Å) was used to account for long-range electrostatic interactions. All other non-bonded interactions were calculated using switching function to smooth interaction energies to zero between 10 and 12 Å. Coordinates were saved every 1 ps for analysis. All MD trajectories were calculated using NAMD 2.9 (Phillips *et al.*, 2005). Trajectory data were analysed using VMD (Humphrey *et al.*, 1996) and customized scripts.

2.2.2.3 Trajectory Analysis

In silico mutations of the SSB-Ct (Ile175Leu (DDDLPF) and Asp172Asn, Asp173Asn, Asp174Asn (NNNIPF) were generated using VMD Mutator Plugin (Humphrey *et al.*, 1996).

As measure of the stability of proteins the radius of gyration (r^2_{gyr}) was calculated according to:

$$r_{gyr}^2(t) = \frac{\Sigma w_i(r_i - \bar{r}_i)^2}{\Sigma w_i} \qquad \textbf{Eq. 2.1}$$

Where r_i is the position of *i* atom from the center of mass.

Hydrogen bonds were analyzed as follows: A hydrogen bond between a donor and an acceptor was considered to exist if the relevant groups satisfied distance (< 3.5 Å) and angle (< 20°) cut-offs. Salt bridges were analysed using the same distance cut-off. These analyses utilized the Hydrogen Bonds and Salt Bridges Plugins in VMD.

VMD Volmap plugin was used to generate weighted atomic density map of SSB-Ct peptide in all simulations (Humphrey *et al.*, 1996). The system of interest is divided into three-dimensional grid (spacing: 1 Å). At each grid point and time-point the atomic density is calculated by replacing each atom with a normalised Gaussian distribution, with 1 standard deviation being equal to its atomic radius. It is used to present each atomic centre and its contribution to atomic density.

The PME algorithm was used in order to calculate long-range electrostatics of the systems (Darden *et al.*, 1993). The ensemble averaged electrostatics calculations were carried out using PME Electrostatic plugin in VMD (Aksimentiev and Schulten, 2005).

The clustering analysis was carried out using the Clustering tool plugin in VMD (Luis Gracia, *unpublished program*). Clusters of SSB-Ct peptide were generated with 3 Å RMSD cut-off within clusters.

2.2.3 Surface plasmon resonance competition assay

SPR measurements utilized a Biacore T200 instrument (GE Healthcare) at 20 °C to measure the competition of compounds for the DnaG/SSB-Ct peptide interaction. The buffer contained 10 mM HEPES (pH 7.4), 3 mM EDTA, 100 mM NaCl, 2% (*v/v*) DMSO, 1 mM DTT and 0.05% (v/v) P20 surfactant (GE Healthcare).

The SSB-Ct peptide immobilization onto the streptavidin chip surface was achieved by gradually increasing the concentration and injection time of N-terminally biotinylated

SSB-Ct peptide [Biotin-(Ahx)-GS**APSNEPPMDFDDDIPF**; where Ahx an amino-hexanoate spacer followed by 16 C-terminal residues of SSB highlighted in bold] to obtain the highest level of immobilization.

The fragments were mixed with the RCD to bring the final concentration of RCD and fragment to 30 μM and 1 mM respectively. Each sample was mixed for approximately 15 min prior to measurements. Mixtures were injected separately to flow cells 1 and 2. Flow cell 1 served as a reference. A flow rate of 5 μL min^{-1} was used during the 60 s injection and 60 s dissociation phases. All experiments were conducted at 20 °C.

2.2.4 NMR Spectroscopy

2.2.4.1 Reference spectra

1D ^{1}H NMR spectra for each individual compound in the absence of protein (1 mM in 50 mM phosphate buffer (pH 7.0), 100% $^{2}H_2O$) were collected at 10 °C on a Bruker-Biospin Avance 600 MHz spectrometer with a cryo-probe and auto sample changer (Bruker-Biospin, Billercia, MA).

2.2.4.2 Saturation transfer difference NMR

STD-NMR experiments to screen fragment libraries were performed in standard NMR tubes using 5 μM unlabelled protein and fragment cocktails in each sample with a concentration of ~ 250 μM for each fragment. The sample volume was 500 μL with 100% $^{2}H_2O$ buffer containing 50 mM phosphate buffer pH (7.8), 50 mM NaCl and 1 mM DTT.

STD-NMR experiments were carried out at 10 °C on 600 MHz Bruker Biospin Avance spectrometer. Saturation of protein was achieved by a 4 s Gaussian pulse sequence train centered at -1 ppm. For reference spectra, a similar saturation pulse was applied 20,000 Hz off-resonance. A 20 ms spin-lock period was applied before acquisition to allow the residual protein signals to decay. The STD dataset were acquired over 64 scans. All NMR data were processed by TOPSPIN 3.1 (Bruker Biospin).

2.2.4.3 2D ^{1}H-^{15}N HSQC NMR

The compounds identified by the SPR and STD screens as hits were confirmed and their binding sites identified by recording ^{15}N-HQSC experiments on uniformly ^{15}N-labelled

DnaGC (100 μM) in the presence of compounds at 3.3 mM dissolved in $^{2}H_6$-DMSO with 50 mM MES (pH 6.0), 60 mM NaCl, 1mM DTT and 3% $^{2}H_2O$ in a final volume of 150 μL. The acquisition time was 30 min for each ^{1}H-^{15}N HSQC experiment with 12 scans. A standard pulse sequence was used for data acquisition. Spectra were recorded at 298 K on a Bruker Avance 600 MHz spectrometer (Bruker-Biospin) equipped with cryoprobe and sample autochanger. The ^{1}H-^{15}N HQSC spectra were processed with TOPSIN 3.2 and analysed using the CCPN suite (Vranken *et al.*, 2005). Compounds were regarded as hits if chemical shift perturbations were observed in the HSQC spectra upon addition of the compound.

2.2.4.4 Chemical Shift Perturbation Calculation

Weighted-average chemical shift perturbation (CSP) was calculated according to Eq. 2.2 (Hajduk *et al.*, 2000).

$$\Delta\delta=\sqrt{[\delta(^{1}H)\text{complex} - \delta(^{1}H)\text{free})^2 + 0.04(\delta(^{15}N)\text{complex} - \delta(^{15}N)\text{free})^2]} \qquad \textbf{Eq. 2.2}.$$

Some minor errors were introduced during peak positioning where peaks in the HSQC were close. These were adjusted manually.

2.2.4.5 Ranking of HSQC Hits

In order to identify the most potent hit compounds in the screen, the hits need to be ranked based on the magnitudes of the chemical shift perturbations. A compound was considered a hit if there was a peak movement, and in cases where the peak shifts were small, the fragment was ranked 1 (weak binder), if the shift was obvious but still within overlap distance of the equivalent apo-protein peak, the hits were ranked 2 (moderate binder), and finally if the peak was shifted away the equivalent apo-protein peak with no overlap, the compound was grouped as rank 3 (strong binder). The disappearance of peaks identifies a compound as being in intermediate exchange (rank 4: strong hit in intermediate exchange).

2.2.4.6 1D ^{19}F NMR Experiments

1D ^{19}F spectra were recorded on a Bruker Avance III 400 MHz NMR spectrometer equipped with the two-channel BBO probe with z-gradient, at standard 300 K sample temperature. All 1D ^{19}F spectra were recorded with 256 scans for fragment and complex samples respectively. Fragment dissolved in $^{2}H_6$-DMSO were diluted in HSQC buffer to give final fragment and protein concentration of 1 mM and 50 μM respectively.

2.2.4.7 K_D and LE Measurements

Binding affinities were estimated by incremental titration of compounds into protein, and recording ^{1}H-^{15}N HSQC experiments at each concentration point. Compound solubilities were tested to determine the highest concentrations of ligands to be used in assays. Equilibrium dissociation constants from NMR titration data were derived using the "single site specific binding with ligand depletion" model in *GraphPad Prism 6.0* (Eq. 2.3)

$$\Delta\delta=\Delta\delta_{max}/2(1+[L_0]/[P_0]+K_D/[P_0]-\sqrt{1+[L_0]/[P_0]+K_D/[P_0]^2})-4[L_0]/[P_0] \quad \textbf{Eq. 2.3}$$

Where the $[L_0]$ and $[P_0]$ are total ligand and protein concentrations, $\Delta\delta_{max}$ is the maximum chemical shift difference observed.

The ligand efficiency was calculated using **Eq. 2.4**:

$$LE = \Delta G/HAC = -RTlnK_D/HAC \quad \textbf{Eq. 2.4}$$

Where ΔG is the Gibbs free energy, HAC is the number of heavy (non-hydrogen) atoms, R is the universal gas constant, T is absolute temperature and K_D is equilibrium dissociation constant.

2.2.4.8 NMR Guided Molecular Docking

NMR data were used to obtaining constraints for docking, using the simple linear relationship considering between the magnitude of the CSP and the distance from the amino acid backbone amide atoms and the nearest ligand atom according to protocol described by Stark *et al.* (Stark and Robert Powers, 2008). AutoDock Tools 1.5.6 (Trott and Olson, 2010) was used to prepare protein and ligand structures. The protonation state of the titratable groups in the protein were assigned at pH 7.0 using PROPKA 3.1 (http://propka.ki.ku.dk) (Li *et al.*, 2005). Polar hydrogen atoms and atom based Gasteiger partial charges were added. Nonpolar hydrogen atoms were merged with the parent atom. The protein structure was taken from previously solved crystal structure of DnaGC with SSB-Ct peptide (Tak Lo, 2012), where protein was treated as rigid body. The docking calculations were performed using AutoDock Vina 1.1.2 (Trott and Olson, 2010). The calculations were done with an "exhaustiveness" of 1024 with a grid points separated by 1.0 Å, and a grid size of 16 × 16 × 14 Å. The ligand-binding residues (those with CSP greater than 0.02 ppm upon fragment binding) were used to generate parameters for NMR-based constrained docking. Constraints were added in these perturbed parts of the protein to identify the most likely binding site for hits (Stark and

Robert Powers, 2008). The grid box was large enough to include the SSB-Ct binding site. Vina generates low-energy binding poses by evaluating the combined energetic contributions of torsion, steric repulsion, hydrogen bonding, and hydrophobic interactions between ligand and protein binding pocket. Ligand data were obtained from the ZINC database in mol2 format (Irwin and Shoichet, 2005).

2.2.4.9 3D ^{15}N-NOESY-HSQC

^{15}N-labelled DnaGC with and without compound was used to record 3D ^{15}N-NOESY-HSQC (τ_m 150 ms) experiment in 5 mm Shigemi tube in the HSQC buffer at 298 K. 440 μM ^{15}N-labelled DnaGC and 1 mM compound was used. Data collection was performed on Bruker 800 MHz spectrometer with cryoprobe.

2.2.4.10 Paramagnetic Sample Preparation and NMR Recording

In order to maintain the solvent accessible cysteine residues in the mutant proteins the samples were treated with a five-fold excess of DTT. The excess of DTT was removed by diluting the sample with NMR buffer and re-concentrating it using Amicon Ultra-4 centrifuge filter with 3 kDa cut-off. The protein was added to a 5-fold excess of aqueous solution of tag in NMR buffer and incubated overnight at room temperature. The excess tag was washed out as described above for excess DTT, and the final sample was transferred to 3 mm Shigemi tube for HSQC acquisition. 1D proton spectrum was recorded for the compound in presence of paramagnetically tagged protein at 1:4 protein ligand ratio. All PCS measurements were done on a Bruker 800 MHz spectrometer equipped with a cryoprobe.

The S-(2,2,5,5-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl)methyl methanesulfonothioate (MTSL) nitroxide spin label was attached to Q445C DnaGC first by remove the DTT as described above, and mixing with 30 equivalents of MTSL dissolved in acetone and incubating for overnight at room temperature. No precipitation was observed in the sample, and excess of MTSL was removed as previously described for the lanthanide tags. The buffer was the same as used in the HSQC NMR experiments. For recording 1D ^{1}H experiments with 1:4 and 1:6 protein-CDS001350 compound, a molar ratio was recorded for 512 scans. 500 mM ascorbic acid was used as reducing agent from the paramagnetic to diamagnetic state [(1-acetyl-2, 2, 5, 5-tetramethyl-3-pyrroline-3-methyl)-methanethiosulfonate] (MTS) for recording the experiment for diamagnetic

reference. Samples were kept at room temperature for an hour for complete reaction. PRE data acquisition was done on Bruker 600 MHz spectrometer with cryoprobe.

2.2.5 Crystallography

Previously published crystallisation conditions for DnaGC [50 mM sodium acetate buffer (pH 4.6), 2.5% PEG 4000, 100 mM $(NH_4)_2SO_4$] (Oakley *et al.*, 2005; Loscha *et al.*, 2004), *E. coli* χ [100 mM NaOAc, 25% (w/v) PEG 4000, 8% (v/v) isopropanol] (Tak Lo, 2012) were used in an attempt to reproduce the high quality crystals for screening. Crystallisation screens were conducted for RCD (JCSG+, PEG, PEG II, PACT) (Newman *et al.*, 2005; Krosky *et al.*, 2006) varying the protein concentration at 4 °C and 22 °C.

Chapter 3

Molecular Dynamic Simulations of SSB-Ct Binding Partners

3.1. Introduction

Molecular dynamic (MD) simulations are widely used in computational chemistry. The method is based on the calculation of the physical forces between atoms and molecules using empirical force fields and finite-difference methods to solve the equations of motion. MD simulations create snapshots of molecular structures as a function of time known as trajectories (Klepeis *et al.*, 2009). Insights can be gained into the mechanisms and behaviours of proteins, nucleic acids, and their complexes in atomic detail (Shaw *et al.*, 2010). The use of MD simulations can lead to new hypotheses concerning molecular function and advance experimental design.

MD simulations are based on the use of Newton's Second Law of motion to calculate the atomic movements. Therefore, atomic masses and a mechanism for calculating the forces on atoms are required. The latter are described by an appropriate empirical force field (**Eq. 3.1**).

$$E_{total} = \overbrace{\sum_{bonds} K_r(r - r_{eq})^2 + \sum_{angles} K_\Theta(\Theta - \Theta_{eq})^2 + \sum_{dihedrals} \frac{V_n}{2}[1 + \cos(\phi - \gamma)]}^{\textbf{Bonded}} + \overbrace{\underbrace{\sum_{i<j}[\frac{R_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}}]}_{\textbf{van der Waals}} + \underbrace{\sum_{i<j}[\frac{q_i q_j}{\epsilon R_{ij}}]}_{\textbf{Electrostatic}}}^{\textbf{Non-Bonded}}$$

Eq. 3.1. Empirical equation for the total energy of a chemical system. Forces can be divided into bonded and non-bonded terms. Modified from (Durrant and McCammon, 2011). The terms r, θ and ϕ refer to bond length, angle, dihedral angles R_{ij} is a distance between i and j atoms.

In biochemistry and molecular modelling, the term "force field" is used to describe both the interatomic potential function and its parameters. For proteins, intensive development efforts have improved the accuracy of simulations (Durrant and McCammon, 2011; Klepeis *et al.*, 2009).

3.1.1. Comparison of SSB-Ct Binding Partners

SSB is a protein interaction hub that binds to at least 14 partners in bacterial cells (**Table 1.3**). Currently it is known that most SSB-binding partners interact with it through the SSB-Ct motif, which is highly conserved among bacterial species (**Figure 3.1**). More recent experimental data from *E. coli* suggest that the C-terminal tail of SSB competes with ssDNA for binding to its own OB-domain (Kozlov *et al.*, 2010; Shishmarev *et al.*, 2014; Su *et al.*, 2014).

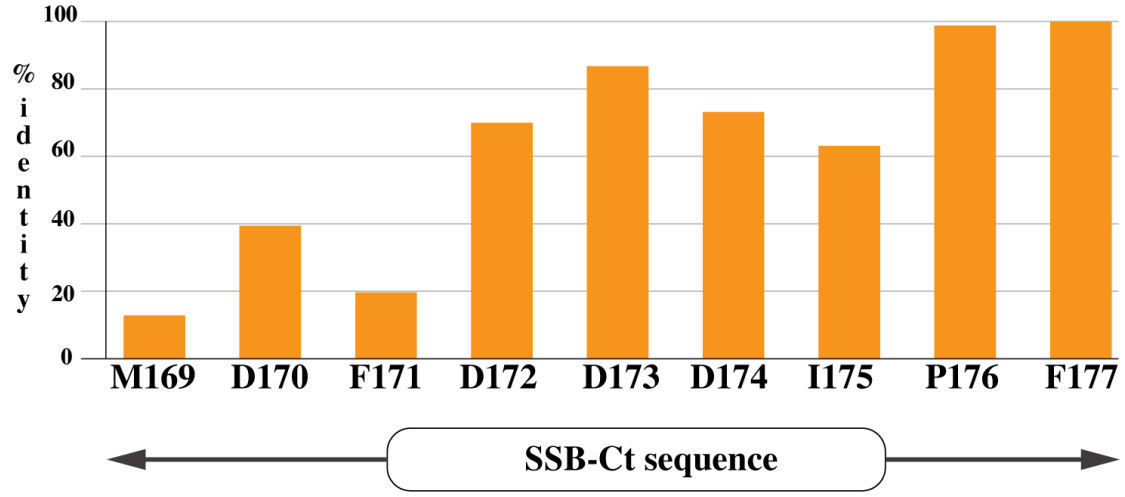


Figure 3.1. Histogram representing the sequence identity of last nine residues of SSB-Ct among 248 bacterial species. Modified from (Lu and Keck, 2008).

Currently, there are four crystal structures reported of proteins in complex with SSB-Ct or similar peptides: DnaGC (Tak Lo, 2012), RecO (Ryzhikov *et al.*, 2011), ExoI (Lu and Keck, 2008) and Pol III χ/ψ subunit (Marceau *et al.*, 2011). Common features can be observed in these SSB-Ct/partner interactions. Where present in the crystal structures, the acidic residues (the "DDD" sub-motif) are not observed in electron density or have high B-factors relative to the residues closest to the C-terminus (the "IPF" sub-motif), which is relatively linear and binds to a hydrophobic pocket on the surfaces of binding partners. The SSB-Ct peptide can therefore be categorized as a SLiM (Short Linear Motif). SLiMs typically contain 3 to 10 amino acid residues who's primary function is to bind a partner molecule (Via *et al.*, 2015). The four crystal structures of complexes with SSB-Ct reveal common features: a hydrophobic pocket that binds to the IPF sub-motif, and this pocket is surrounded by basic residues (**Figure 3.2**).

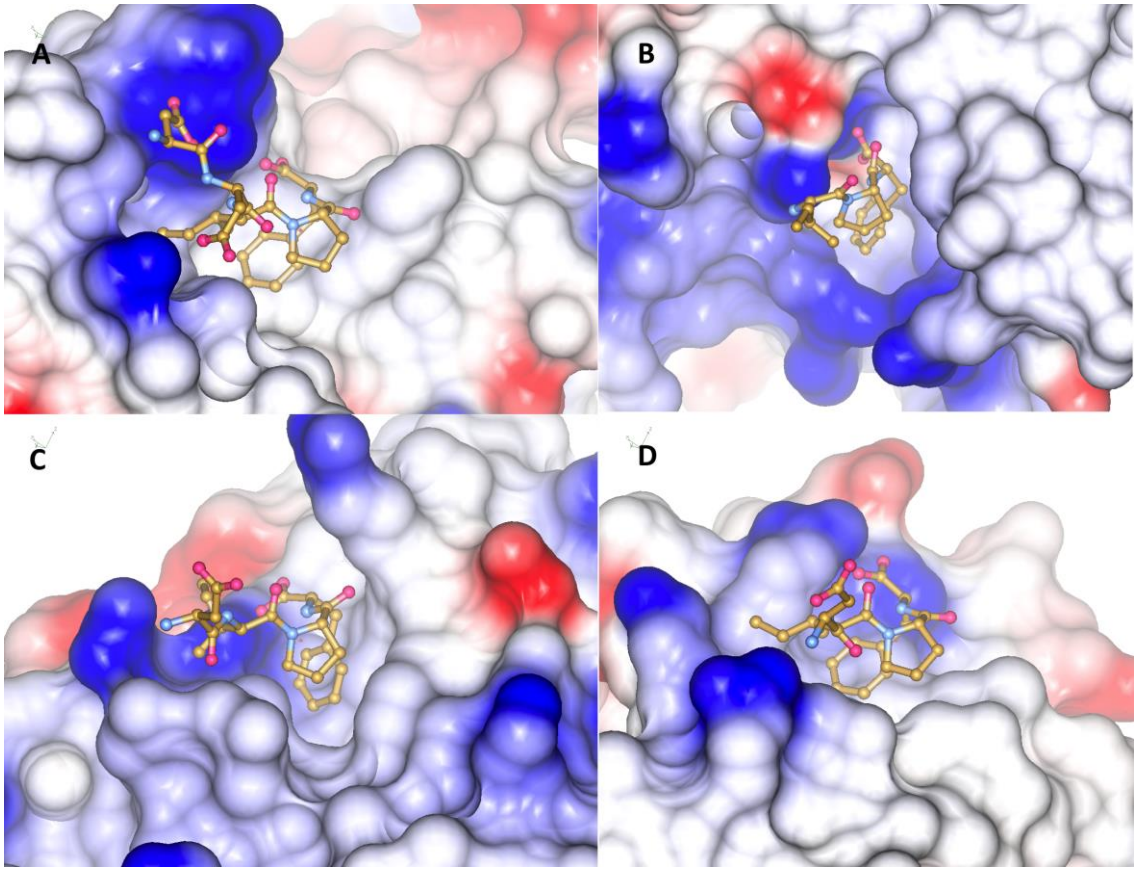


Figure 3.2. Molecular surface of DnaGC (A), RecO (B), ExoI (C) and Pol III χ/ψ (D) with SSB-Ct shown in stick form. Structures are shown with all SSB-Ct peptides in approximately the same orientation, Positive and negative electrostatic potentials are shown in blue and red respectively, Figure created with Coot (Emsley and Cowtan, 2004).

In the case of DnaGC, the binding pocket accommodating the IPF sub-motif of SSB-Ct is lined by Leu446, Met451, Ile455, Pro480, Leu484, Thr515 and Leu519 (**Figure 3.3**). The C-terminal carboxylic acid of SSB-Ct (Phe177) forms a salt bridge with the guanidinium and amino groups of Arg452 and Lys447. The Ile175 residue forms hydrophobic contact with Leu446 and Leu519.

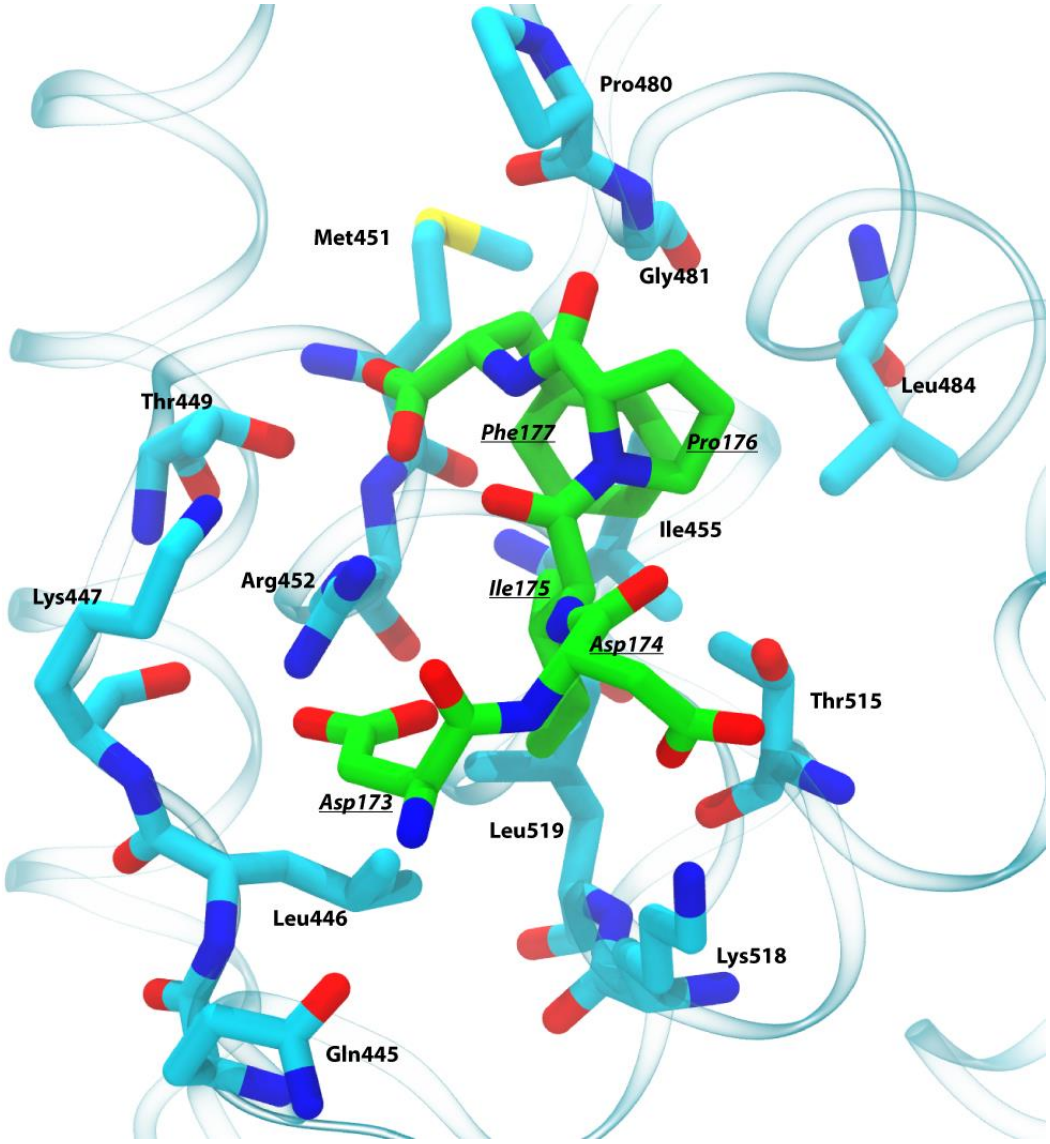


Figure 3.3. Crystal structure of SSB-Ct in complex with DnaGC/SSB-Ct (Tak Lo, 2012). SSB-Ct peptide residues are underlined. Generated by VMD (Humphrey et al., 1996).

Asp174 makes a hydrogen bond and salt bridge with the side chains of Thr515 and Lys518 respectively. Asp173 forms salt bridge with Arg452 and Lys447. Mutation of Lys447 and Lys518 residues to alanine resulted in a 10-fold decrease in K_D, while the Arg452Ala mutant eliminated the interaction (Tak Lo, 2012; Naue *et al.*, 2013). An alignment of SSB-Ct peptide in all four structures reveals that the conformation of the IPF sub-motif is surprisingly similar (**Figure 3.4**).

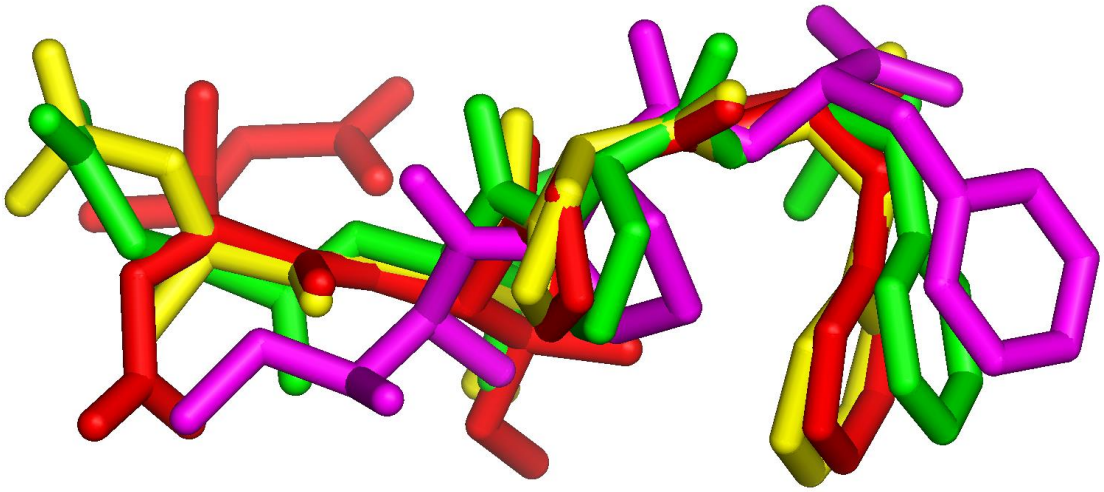


Figure 3.4. Structural alignment of SSB-Ct peptide in complex with DnaGC (red), RecO (magenta), ExoI (green) and χ (yellow). Image generated by PyMol Molecular Graphics System, Version 1.3 Schrodinger, LLC.

The RMSD of all non-hydrogen atoms in the IPF sub-motif as observed bound to ExoI, RecO and χ/ψ are 1.7 Å, 2 Å and 0.5 Å with respect to the IPF sub-motif bound to DnaGC. In all four complexes considered here, an arginine guanidinium group (Arg452 in DnaGC) forms a salt bridge with the terminal carboxylic acid group of SSB-Phe177. In contrast to the relatively well-conserved conformation of IPF sub-motif, the conformations of the acidic residues (the DDD sub-motif), where visible, appear highly variable. It should be noted that in the crystal structure of the Pol III χ/ψ complex, a truncated peptide (WDIPF) was used instead of the SSB-Ct peptide.

3.1.2 Aim

Two hypotheses regarding the interaction of SSB-Ct with its binding partners may be advanced based on the crystal structures: (1) the IPF sub-motif is relatively rigid and makes well defined interactions with its binding partners, and (2) the DDD sub-motif is relatively disordered and engages in non-specific electrostatic interactions with basic residues surrounding the immediate vicinity of the SSB-Ct binding pocket. To further illuminate these interactions, the dynamics of the interactions between SSB-Ct and its binding partners were probed by molecular dynamic simulations. This was anticipated to reveal conserved interaction patterns, the basis for the diversity of conformations observed for the three acidic residues, and to clarify the role (if any) of crystal contacts on peptide conformation. Furthermore, *in silico* examination of the impact of the

following mutations in SSB-Ct (Asp172Asn, Asp173Asn, Asp174Asn (NNNIPF) and Ile175Leu (DDDLPF)) bound to DnaGC were examined.

3.2 Results and Discussion

3.2.1 Global Structure Behaviour

The structural fluctuations of each system in terms of RMSD is shown in **Figure 3.5** and *Appendix A*. The RMSDs are generally around 2-3 Å, with no significant changes observed for any of the systems. In contrast, the SSB-Ct peptide shows greater fluctuations (**Figure 3.5B** and *Appendix A*).

The RMSDs for DnaGC/SSB and complexes of DnaGC with mutated SSB-Ct were broadly similar (**Figure 3.5A**). DDDLPF shows greater fluctuations; on average 3.8 Å RMSD with an increase up to 4 Å after 40 ns, most likely caused by the introduced I175L mutation. The NNNIPF mutant did not show much deviation from the wild type system.

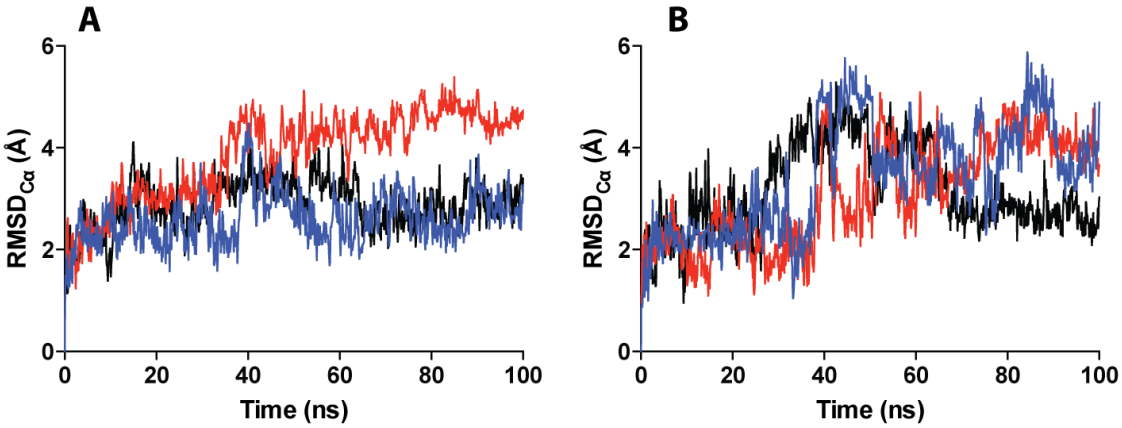


Figure 3.5. Backbone RMSD plotted for each individual chain of DnaGC/SSB-Ct system along the 100 ns MD simulations. (A) represents RMSDs for the entire DnaGC/wtSSB-Ct (black), DnaG/DDDLPF (red), and DnaG/NNNIPF (blue) mutated systems, (B) is RMSD for the SSB-Ct peptides. Black is DnaGC/wtSSB, red is DDDLPF and blue NNNIPF system respectively.

To monitor the structural stability of the systems, the radius of gyration (R_g an indicative of compactness of the systems) was calculated for all trajectories. In all cases, the radius of gyration did not show any significant changes from its initial value. All systems remained stable during the 100 ns simulation. In addition, the secondary structures of the complexes did not show any significant changes during the simulations (*data not shown*).

3.2.2 Dynamical Behaviour of SSB-Ct

Detailed examination of all simulations showed structural similarities and similar patterns of interactions. The available crystal structures of all four complexes are missing some or all acidic residues of the SSB-Ct peptide, suggesting high mobility. The residue-by-residue RMSD (rRMSD) for each Cα atom was calculated from molecular dynamic simulations (**Figure 3.6**). For each trajectory, the SSB-binding partners were superimposed in their starting positions prior to calculations.

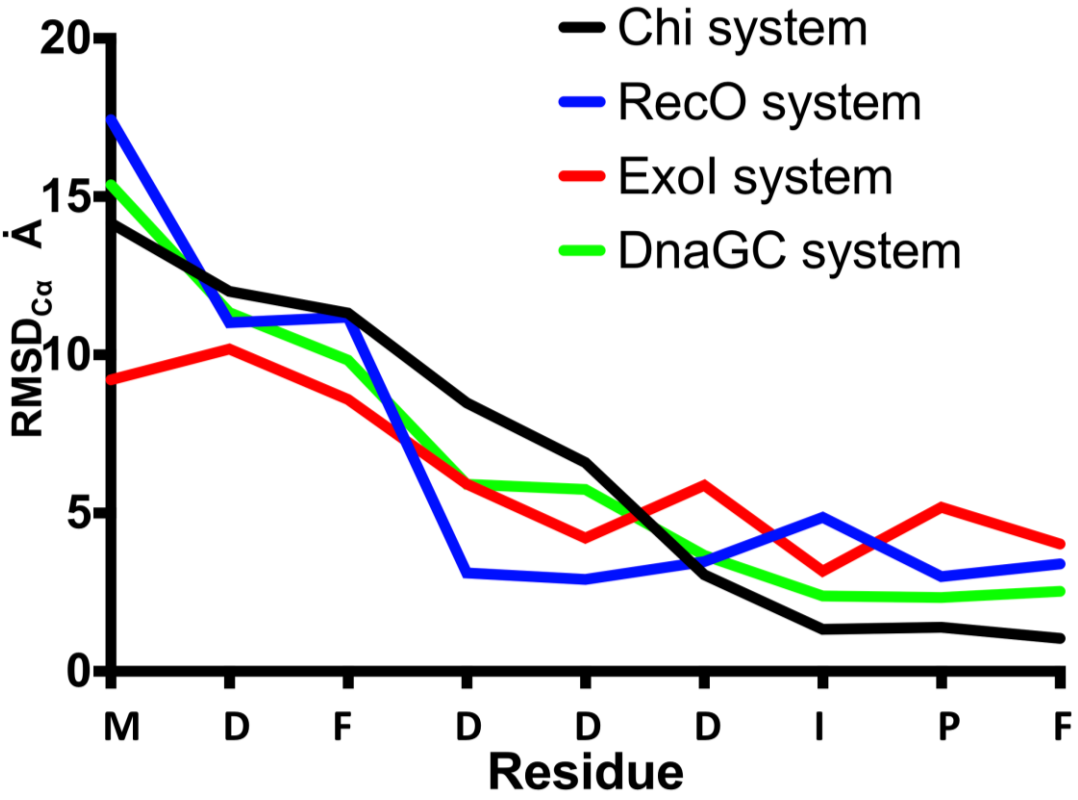


Figure 3.6. RMSD for SSB-Ct Cα atoms for all four systems over 100 ns.

In general, the SSB-Ct peptides showed fluctuations that increased with distance from the C-terminus. SSB-Ct residues not in direct contact with the binding partner sample a greater diversity of conformations. The order parameter (S^2) was calculated for the backbone amide nitrogen atoms of SSB-Ct peptide in all four complexes (**Figure 3.7**) as follows:

$$S^2 = \frac{1}{2}[3\sum_{i=1}^{3}\sum_{j=1}^{3}\langle\mu_i\mu_j\rangle^2 - 1]$$

where μ_i and μ_j are the coordinates of the normalized backbone NH vectors, decomposed into three spatial dimensions (i, j=1,2,3). Values range from 0 (perfect disorder) to 1 (perfect order).

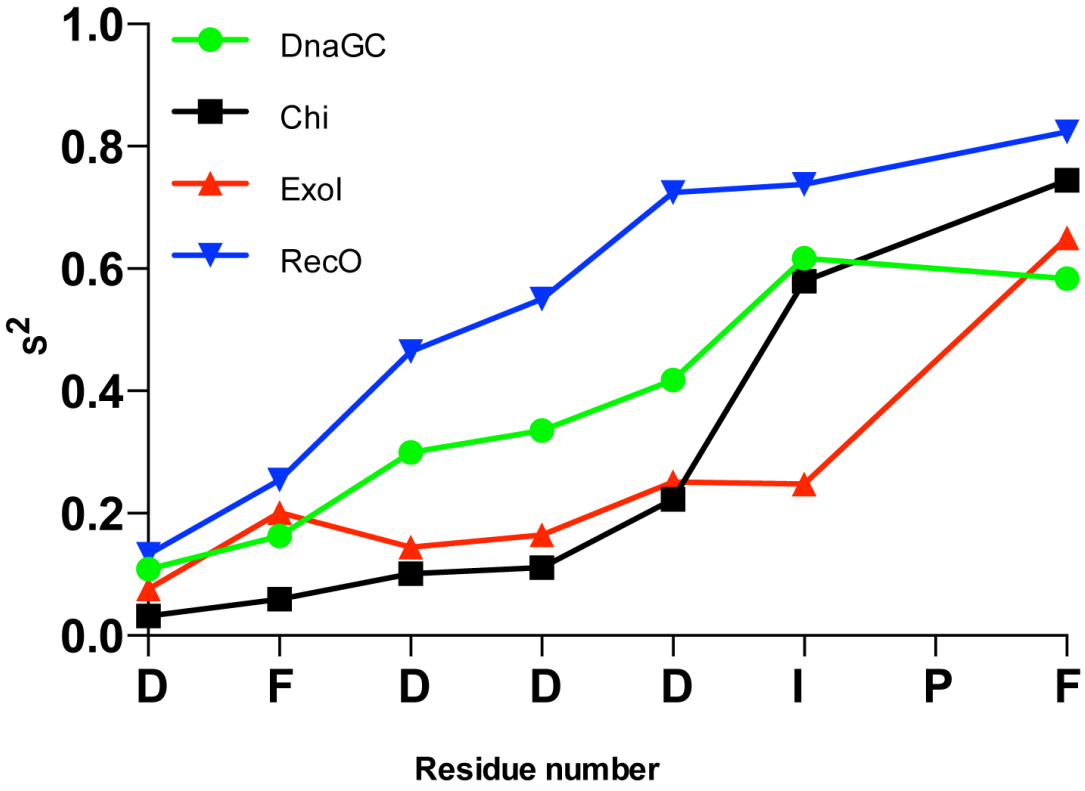


Figure 3.7. Order parameter (S^2) for the SSB-Ct peptide in all four simulations.

The general trend is in agreement with rRMSD calculations, showing that the IPF sub-motif has higher order (S^2 closer to 1) compare to the rest of the peptide.

The volume of space explored by SSB-Ct was visualized by calculating weighted atomic density maps for the four systems (**Figure 3.8**).

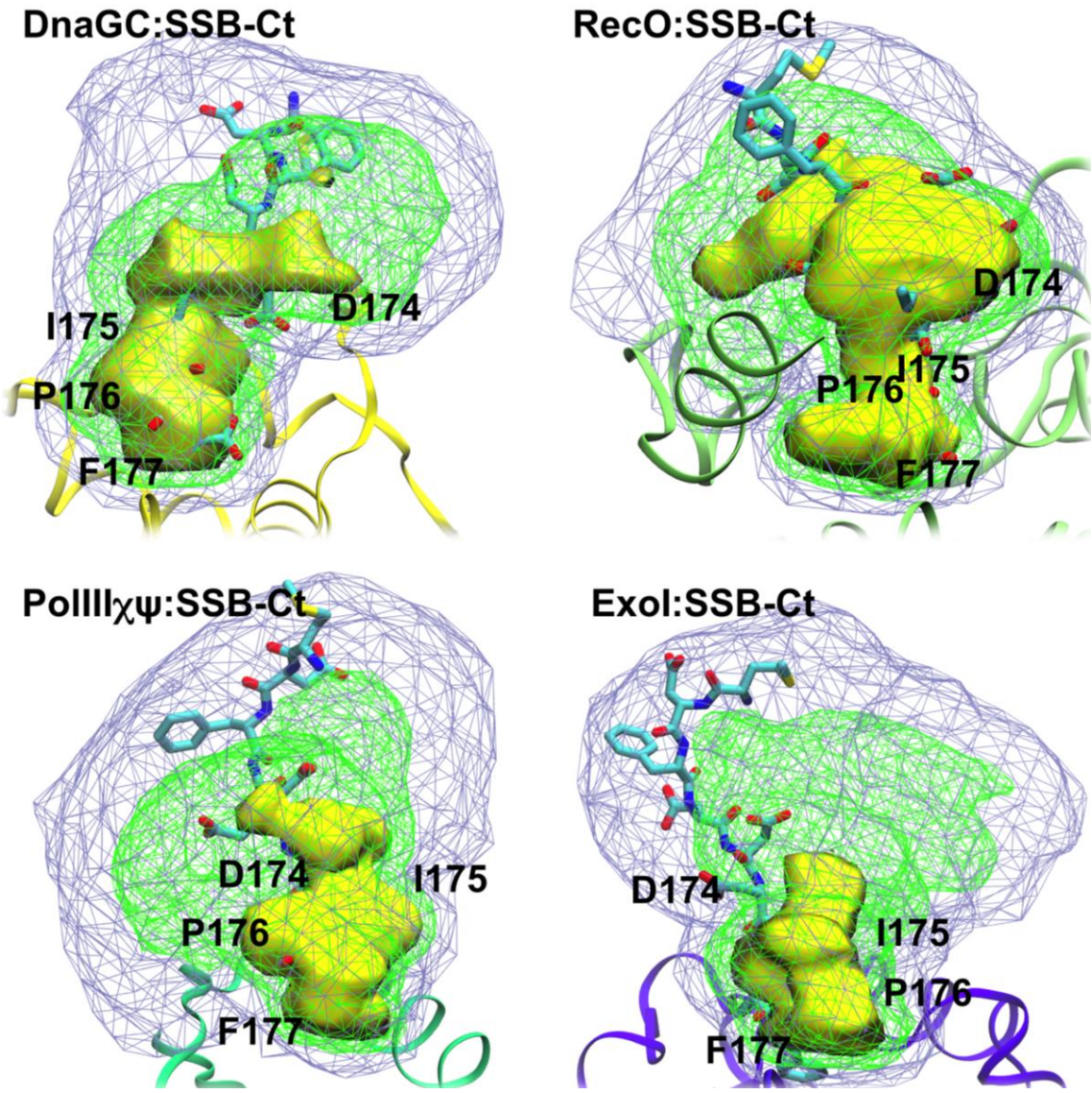


Figure 3.8. The weighted atomic density of SSB-Ct peptides in the four simulated complexes, sampled at 1 ns intervals and averaged over 100 ns. The yellow surface, green wireframe and blue wireframe indicate density contours of 0.07×10^{-6} Da Å $^{-3}$, 0.0017 Da Å $^{-3}$ and 0.0014 Da Å $^{-3}$ respectively. Density was calculated using Gaussian distribution of width equal to the atomic radii.

The weighted atomic density analysis shows that the IPF sub-motif remains buried in the binding pocket while the proceeding residues explore a large volume of space.

In all simulations, the salt bridge formed between an arginine residue and the α-carboxyl group of Phe177 in the SSB-Ct motif was maintained. This is congruent with the observation that mutation of the salt-bridge-forming arginine residue to alanine abolishes interactions in all systems (Lu and Keck, 2008; Marceau *et al.*, 2011; Naue *et al.*, 2011; Naue *et al.*, 2013; Tak Lo, 2012). Curiously, no hydrogen bonds are formed between the amide group linking the proline and phenylalanine in SSB-Ct and any of the binding partners considered here.

Clustering analysis was carried out for SSB-Ct peptide in the four wild type system trajectories, as shown in **Figure 3.9**.

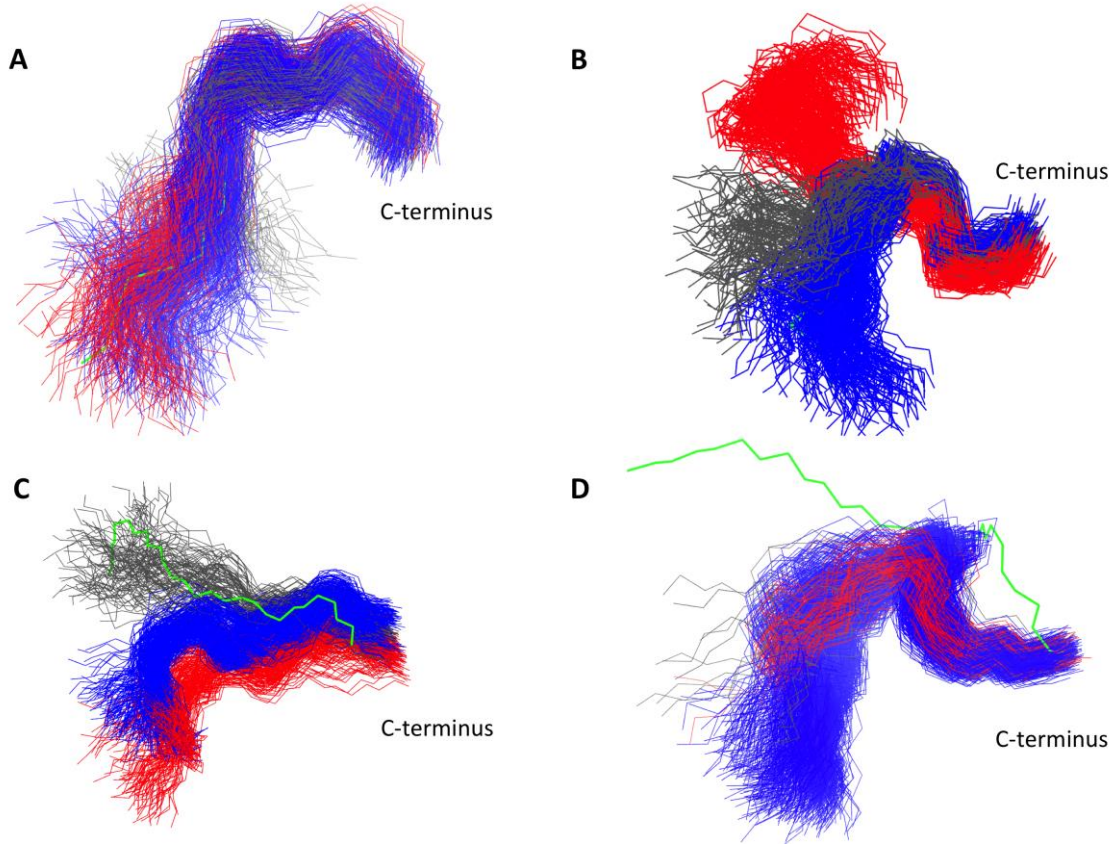


Figure 3.9. Three most populated clusters SSB-Ct conformations bound to DnaGC (A), χ/ψ (B), ExoI (C) and RecO (D). The starting orientation of the peptide is in green, where blue, red and grey are the first, second and third most populated clusters respectively.

Clustering analysis shows that SSB-Ct conformations cluster relatively tightly in the DnaGC and RecO simulations, and that structurally more distinct clusters occur in the χ/ψ and ExoI simulations. This could indicate that a dominant binding mode exists for the DnaGC and RecO. In the case of DnaGC, this tight clustering can be explained by salt-bridge interactions of Arg452 and Lys447 with the central aspartate residue of the DDD sub-motif (Asp173). A similar pattern occurs in RecO, where a cluster of arginine residues (Arg210, Arg225 and Arg229) engage in salt bridge interactions with the residues in the DDD sub-motif. In the case of χ/ψ and ExoI, the peptide clusters represent relatively low-population conformations where acidic residues form salt bridges with residues on the surface of those proteins. The patterns are detailed in **Table 3.1**.

The backbone dihedral angles of SSB-Ct peptide (DnaGC/wtSSB) were monitored during the simulation. Residue Pro176 likely plays the role of a "conformational lock"

orienting the residues on either sides of it. The cyclic pyrrolidine group of Pro176 fixes the backbone dihedral Φ angle (C-N-Cα-C) in the range preferred for α helices (~60°), but leaves dihedral angle the ψ (N-Cα-C-N) relatively unrestrained. In the DnaGC complex, Φ = -85° and ψ = 37°.

It was observed that Phe177 in DnaGC/wtSSB-Ct complex showed significant structural differences compared to the *in silico* Ile175Leu mutant. In the DnaGC/DDDLPF simulation, Phe177 moves away from the pocket, and at 60 ns it reaches a critical point where its phenyl ring is no longer in the pocket, which explains the increase in RMSD after 60 ns (**Figure 3.5**). *In silico* mutations of SSB-Ct peptide (Asp172Asn, Asp173Asn and Asp174Asn) had a significant impact on orientation of peptide bound to DnaGC; the loss of electrostatic interactions appeared to allow movement of Phe177 out of the pocket. The salt bridge between the Phe177 α-carboxyl group and Arg452 breaks at about 84 ns mark, which makes peptide more flexible and completely changes its orientation. Therefore, while the DDD sub-motif does not participate in long-lived interactions on the surface of SSB-Ct, it does appear to help stabilize the binding of the IPF sub-motif.

In general, interactions observed in available crystal structures were maintained in the MD simulations (**Table 3.1**). The SSBCt-Phe177 was observed to engage in salt bridge interactions with arginine with a population > 50% (**Table 3.1**). Consistent with the lack of hydrogen bonds in the crystal structures, the amide group linking SSB-Ct residues Pro176 and Ile175 form hydrogen bond with low occupancies (**Table 3.1**).

The DDDLPF mutant has reduced populations of some of the hydrogen bonds. In the case of the RecO/DDDLPF system, the Phe177:Arg132 interaction showed decreased hydrogen bond population (45%) compared to wild type (69%). Nevertheless in DnaGC/DDDLPF Pro176 gets closer to Lys447 to form short-lived hydrogen bond.

Of the four aspartate residues in SSB-Ct, Asp173 engages in the most direct hydrogen bonding or salt-bridge interactions (**Table 3.1**). Asp to Asn mutations in SSB-Ct dramatically decreased the population of direct interactions, suggesting that the presence of formal charges are important for the SSB-Ct/binding partners interactions.

Table 3.1. Occupancies of hydrogen bonds between SSB-Ct residues and all four wild type systems. Interactions with less populations of > 5% were excluded.

	D170		F171		D/N172		D/N173		D/N174		I/L175		P176		F177	
DnaGC	-	-	-	-	-	-	Gln445	10-20%	-	-	-	-	Lys206	14.3%	Arg452	79%
	-	-	-	-	-	-	Lys447		-	-	-	-	-	-	-	-
	-	-	-	-	-	-	Arg448		-	-	-	-	-	-	-	-
DnaGC-DDDLPF	-	-	Lys518	8%	-	-	Arg448	19%	Lys518	11%	-	-	-	-	Thr449	10%
							Arg452	~100%							Arg452	60%
															Lys447	10%
DnaGC-NNNIPF	-	-	-	-	-	-	Gln445	15%		-	-	-	-	-	Arg452	69%
							Pro444	6%								
ExoI	Arg327	78%	-	-	-	-	-	-	-		-	-	-	-	Arg148	32%
															Arg203	83%
RecO	Arg225	45%			Arg203	2%	Arg207	68%	Arg203	32%	-	-	-	-	Arg132	69%
	Arg229	10%			Arg229	15%	Arg210	53%							Lys206	15%
RecO-DDDLPF	-	-	-	-	-	-	Arg229	15%	Arg210	11%	-	-	Lys206	11.2%	Arg225	40%
															Arg132	45%
									Arg203	62%					Arg210	46%
															Arg225	31%
															Lys206	6%
χ	-	-	-	-	-	-	Arg135	13%	Arg135	65%	-	-	-	-	Arg148	69%

3.2.3 Electrostatic Interactions

The importance of acidic residues in SSB-Ct, and the presence of basic residues around the SSB-Ct binding site was noted above. To understand the relationship between the architecture of the complexes and electrostatic forces at play, the electrostatic potential has been mapped (**Figure 3.10**). The contrast between the wild type SSB-Ct and NNNIPF mutant can be clearly seen in the difference between their electrostatic potentials: the region of negative potential due to the SSB-Ct peptide is juxtaposed with the positive potential around the SSB-Ct binding pocket. In NNNIPF-Ct peptide, only intact Asp170 was involved in charged interactions with Lys447 and Lys518 (**Table 3.2**). These data support the hypothesis that non-specific, long-range electrostatic interactions contribute to SSB-Ct/binding partner interactions.

Tab 3.2. The list of residues of DnaGC involved in salt bridge formation with wild type, DDDLPF and NNNIPF mutated complexes over the trajectories.

SSB-Ct residues	**wtSSB-Ct complex**	**DDDLPF complex**	**NNNIPF complex**
Asp170	Lys447, Lys518	Lys447, Arg448, Lys518	Lys447, Lys518
Asp/Asn172	Lys447, Arg452, Lys518	Arg448	-
Asp/Asn173	Lys447, Arg452	Lys447, Arg448, Arg452	-
Asp/Asn174	Lys518	Arg448, Arg452, Lys518	-

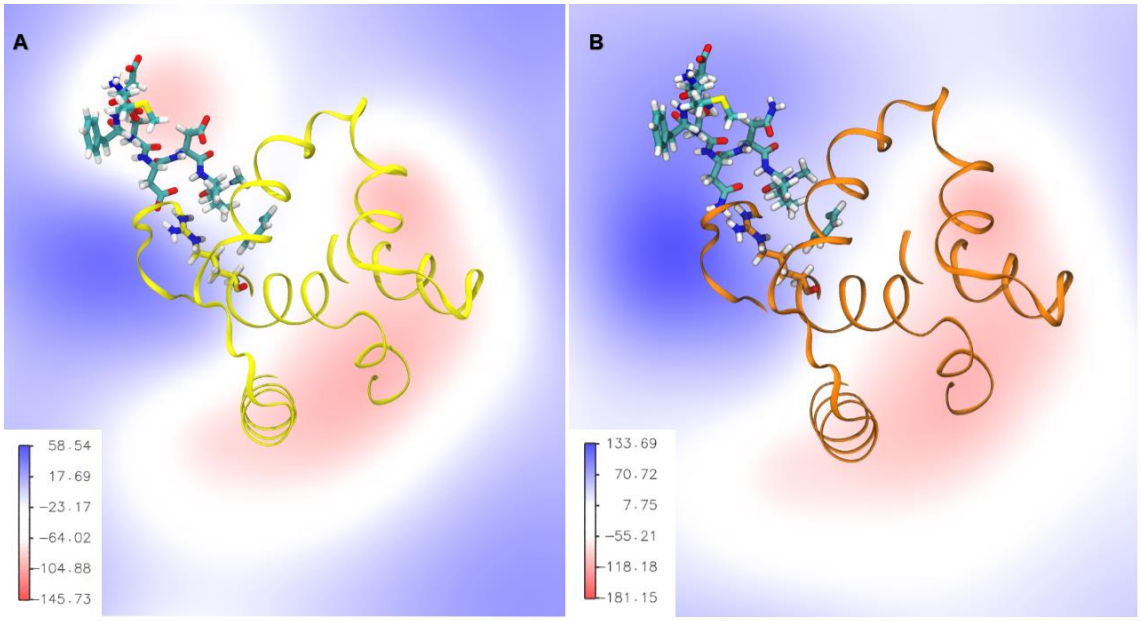


Figure 3.10. Electrostatic potential map orthogonal to the Y axis for wild type DnaG/SSB-Ct (A) and NNNIPF (B) systems. Protein represented in yellow and orange ribbon, the peptide is in stick representation (B). Values are in kT/e.

3.4 Conclusion

A simulation-based approach was used to study the behaviour of SSB-Ct/binding partner interactions. The integrity of the simulated systems was demonstrated by the maintenance of interactions observed in the crystal structures being maintained throughout the simulations. Low-population interactions between SSB-Ct and binding partners were detected that were not observed in the crystal structures (e.g. hydrogen bonding in RecO/SSB-Ct between Arg207 and Aps173; Arg210 and Asp173). These interactions could be further probed by site-directed mutagenesis and binding assays.

The importance of acidic residues in the SSB-Ct peptide was conformed by *in silico* mutagenesis: the loss of charged aspartate residues in the DDD sub-motif destabilized peptide binding. Those data, and the data from the other SSB-Ct complexes suggest that non-specific electrostatic interactions play an important role in SSB-Ct binding. The substitution of Ile175 to leucine was shown to affect the orientation of the Phe177 in the binding pocket, most likely due to steric reasons, which explains the reduction of the measured binding affinities (unpublished data).

Overall, the MD results are in agreement with currently available mutagenesis data supporting that presence of acidic residues in the SSB-Ct are essential for this interaction. Similarities in amino acid composition of binding pockets were determined.

Chapter 4

Fragment-Based Screening Against C-terminal Domain of DnaG primase

4.1.1 Introduction

Screening a fragment library containing hundreds of compounds requires high-throughput capability coupled with the sensitive detection of fragment binding. SPR and ligand-based NMR methods are able to measure weak protein-fragment interactions and are the most popular primary screening tools (B J Davis and Erlanson, 2013). SPR is a powerful label-free biophysical technique for the detection of protein-protein and protein-small molecule interactions in real time (Neumann *et al.*, 2007). To detect binding, receptor molecules are immobilized on the sensor surface, and a solution of analyte (e.g. a fragment) is continuously injected onto the flow cell. The detection of binding is based on SPR phenomenon. Changes in micro-environment on the chip surface affects the reflection of the polarised light angle upon interaction, which is recorded by a detector. The method can be used for competition experiments, where the receptor is immobilized on the surface and the known binding partner is injected with the fragments to be tested. Competition assays provides additional information about the fragment binding location. In addition to measuring binding affinities, SPR can be used to measure binding kinetics (Cooper, 2002) (See section *1.2.2.3*).

NMR provides robust methods for testing the binding of fragments and measuring the binding affinities. Since NMR was first used as a screening tool (Hajduk, 2006), the available methods have evolved. From a screening prospective, NMR-based methods can be divided into two groups: protein- and ligand-based (Dalvit and Stockman, 2002). STD-NMR is the most commonly used ligand-based NMR method. It is composed of *on*- and *off-resonance* experiments. In the *on-resonance* step, selective irradiation of protein protons ensures that saturation is efficiently propagated across the entire protein. In case of ligand binding, magnetization is transferred from protein to bound ligand *via* spin diffusion (Mayer and Meyer, 1999). Care must be taken to choose the range of spectrum for irradiation so as to avoid compound resonances that could result in false positive hits. For that purpose, the -1 to 1 ppm range is used, which is usually free of ligand signals, and is dominated by protein methyl groups. In the event of ligand binding and where protein-ligand ^{1}H-^{1}H distances are short (< 6 Å), magnetization transfers from protein to the transiently bound ligand *via* the intermolecular NOE. Small molecules bound to proteins reflect the properties of the larger molecule, i.e. they exhibit negative NOEs. Compound peaks are therefore decreased in *on-resonance* spectra. In the *off-resonance* experiment, selective pulses are applied to the spectral

region devoid of either protein and ligand signals, e.g. -40 ppm. This generates a reference spectrum, whereby the signals of neither protein nor ligand are saturated. By subtracting the *on-resonance* (saturated) from *off-resonance* (non-saturated) spectra, the final STD spectrum is generated. It contain peaks for compounds that bind to the protein (Mayer and Meyer, 1999; Angulo and Nieto, 2011; Viegas *et al.*, 2011). The STD experiment is illustrated in **Figure 4.1**.

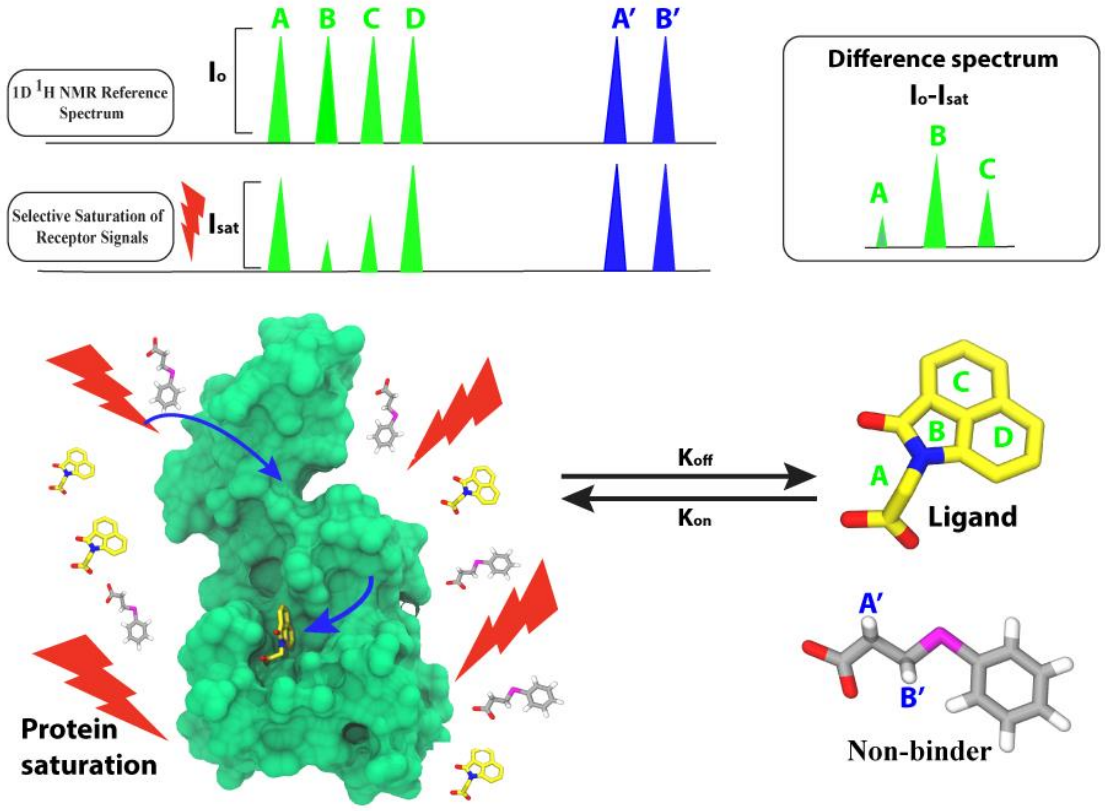


Figure 4.1. Illustration of the STD-NMR experiment. At the top are NMR spectra before and after selective saturation of protein, with final STD spectrum generated (right top), the peaks in green due to a compound that binds to protein from the cocktail, while the peaks in blue correspond to a non-binding compound. At bottom-left is the fragment cocktail with protein. The blue arrows indicate selective irradiation to saturate the protein and its transfer to bound ligand.

Recently, fragment screening for inhibitors of protein-protein interactions (PPI) using ligand-based NMR methods was demonstrated. It was shown that varying the protein concentration (maintaining the ligand concentration fixed at 1 mM), may lead to improved ability to detect weak binders (Dias *et al.*, 2014). To enhance the signal to noise ratio (S/N) or saturation efficiency, protein stability and compound solubility, experiments are often carried out at around 10 °C in 100% ^2H_2O buffer (Begley *et al.*, 2013; Mayer and James, 2002). The STD method is frequently used to screen fragment cocktails. One limitation of fragment screening is the solubility of individual fragments as well as cocktails of fragments in aqueous conditions used for screening. Therefore,

prior to any measurements, the solubility and aggregation properties of ligands in cocktails should be assessed so as to avoid promiscuous binders. Peaks in the 1D ^{1}H-NMR spectra of cocktails should not overlap (LaPlante, Carson, *et al.*, 2013; LaPlante, Aubry, *et al.*, 2013; Doak *et al.*, 2013).

Because of its high-throughput nature, STD-NMR is useful in the early stages of screening. It can be used for deducing the binding affinities and ligand epitope mapping. By monitoring the intensity changes during the experiment, the functional groups in ligands that are in closer proximity to the target protein can be identified (Mayer and Meyer, 2001). Additionally, the principals of STD can be easily transferred to 2D-NMR experiments, such as 2D STD-COSY (Wagstaff *et al.*, 2010). Affinity measurements of weakly binding compounds using STD titration are not accurate due to errors in saturation time, protein concentration and intensity of the monitored signals (Angulo *et al.*, 2010).

In spite of the aforementioned advantages, STD-NMR has a high false positive rate due to the inability to differentiate between specific and non specific-binding, and the aggregation of fragments (Vom *et al.*, 2013; Doak *et al.*, 2013). Therefore, cross validation of hits from STD-NMR screens with other ligand-based methods is recommended (Dias *et al.*, 2014).

4.1.2 Aim

The overall aim was to identify first generation of fragment hits targeting the SSB-Ct binding pocket and catalytic domain of DnaG primase. Initially, an attempt to use X-ray crystallography as the screening method was undertaken, however, suitable crystals were not forthcoming. A SPR competition assay and ligand-based STD NMR were used instead (initially against the RCD). SPR hits were also cross-validated by STD NMR.

4.2 Results and Discussion

4.2.1 Crystallography

Initially, an attempt was made to use protein crystallography as a primary fragment screening method. Purified DnaGC was subjected to crystallization experiments as reported previously (Oakley *et al.*, 2005; Loscha *et al.*, 2004), resulting only in protein precipitation. Therefore, screening for new conditions using the JCSG+ screen

(Newman *et al.*, 2005) was attempted in order to find new conditions and no crystals were observed after several weeks.

Crystallisation trials were then carried out on RCD, varying the protein concentration, temperature (4 °C and 22 °C), using JCSG+, PEG, PEG II and PACT crystallisation suits (Newman *et al.*, 2005; Krosky *et al.*, 2006). Protein precipitated in about half of the conditions, confirming that protein concentration was not low. No crystals were observed even after several months. These could be influenced by presence of flexible linker between two domains of RCD protein, as well as by many other factors affecting crystallisation process (protein batch, temperature, concentration, pH etc.).

Alternative screening approaches, utilising SPR and NMR methods were undertaken.

4.2.2 Primary screening using SPR competition assay

A SPR competition assay was used to test the "first pass screen", a fragment screen by Zenobia Therapeutics. The RCD component of DnaG was used for fragment screening. The primary target of interest was SSB-Ct and DnaG interaction, which has been shown to occur *via* C-terminal domain of DnaG primase (Tak Lo, 2012; Naue *et al.*, 2013). The presence of two domains in the RCD could increase the hit rate, as the presence of two domains provides additional surface area for fragment binding. However, this would require effects from fragment binding to the RNAP domain to be transmitted by some allosteric mechanism to the C-terminal domain. The SPR competition assay with immobilised SSB-Ct peptide on the chip surface (**Figure 4.2A**) and RCD pre-incubated with fragments as the analyte was conducted. An experimental setup using the biotinylated SSB-Ct peptide immobilized on the chip surface and using RCD as an analyte resulted in a relatively large mass change on the chip surface and hence a strong SPR signal. The solubility of fragments under experimental conditions was tested at concentrations of 1 mM. Solutions containing fragments were flowed across the chip surface prior to peptide immobilization to eliminate those that bound non-specifically to the gold surface of the chip and which would give a false positive response. It was observed that the level of response showed a gradual decrease compared to initial levels, most likely due to loss of peptide from the surface.

Subsequent injections of RCD and RCD-fragment mixtures, generated sensorgrams such as the one shown in **Figure 4.2B**. As expected binding affinity is weak (with K_Ds in the millimolar range), small changes (in the order of 5-10 RU) are considered positive hits (**Figure 4.2B**).

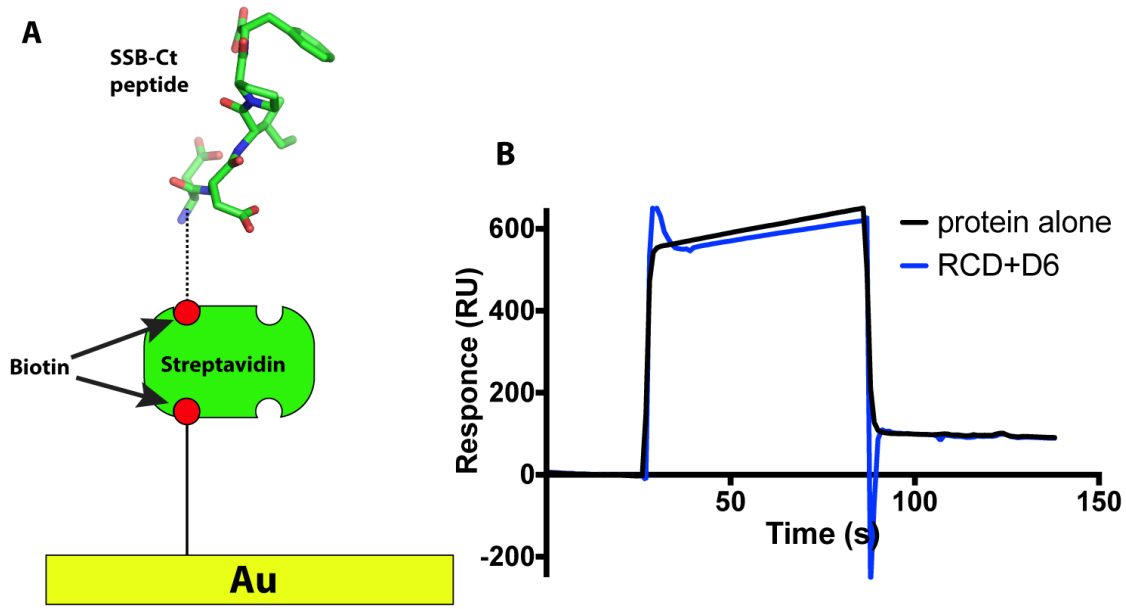


Figure 4.2. SPR experimental setup. (A) Biotinylalted SSB-Ct peptide immobilized on streptavidine chip surface. (B) competition sensorgram for the compound D6. The black trace corresponds to RCD and blue with RCD + D6.

The "first pass screen" using the SPR competition assay identified 6 compounds that compete with SSB-Ct for the RCD. The structure of all of the six fragments are shown in **Figure 4.3**.

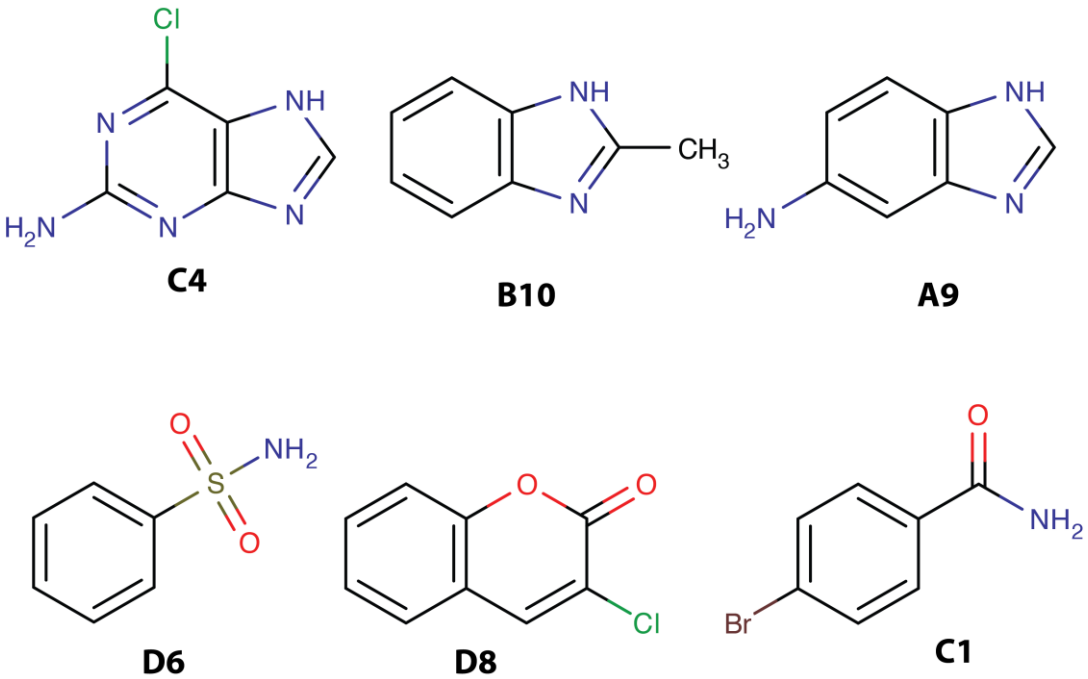


Figure 4.3. The chemical structures of SPR hits.

4.2.3 Validation of SPR hits and identification a new hits by STD-NMR

The 6 fragments from the "first pass screen" were retested by STD-NMR. Each fragment was tested separately at 1 mM concentration. STD gave rise to smaller peaks in the NMR spectra for all of 6 compounds confirming the binding to the RCD protein (**Figure 4.4)**.

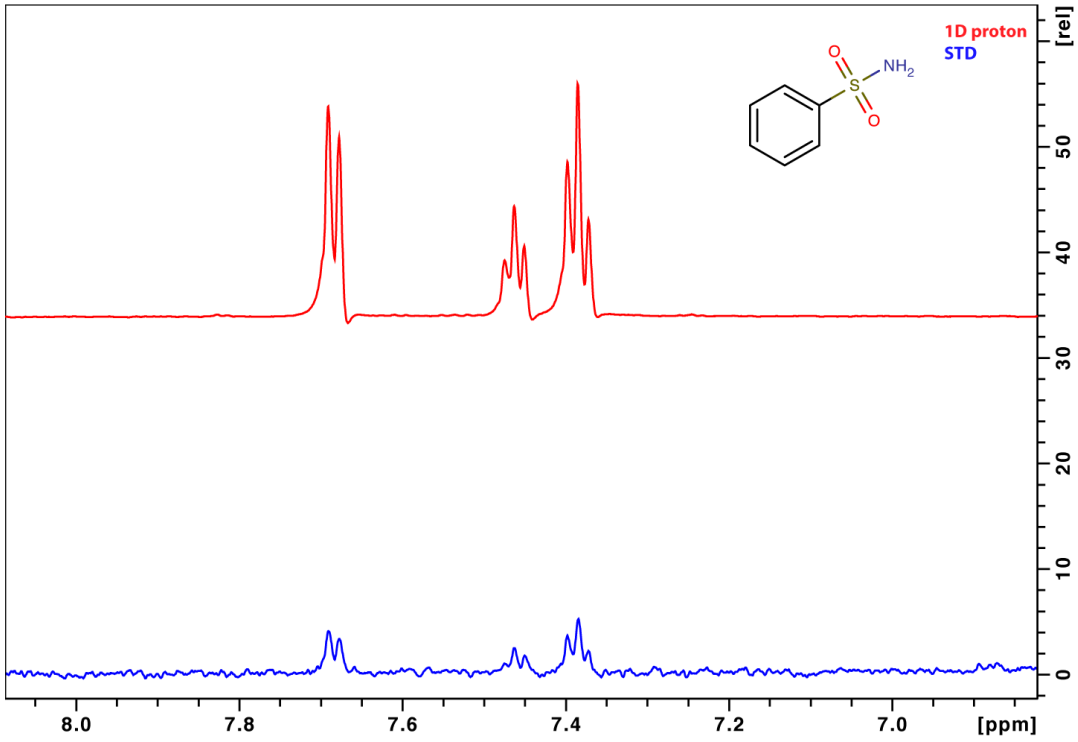


Figure 4.4. STD spectrum of SPR hit D6 confirmed by STD-NMR. Red represents 1D ^{1}H of ligand alone and blue is final STD spectrum

Concurrent with the confirmation of SPR hits by STD-NMR, the MIPS library was screened. The MIPS library was designed specifically for NMR screening and contains about 1200 fragment with diverse structures (Doak *et al.*, 2013). Each cocktail contained 6 fragments, with diverse functional groups chosen so as to minimize overlap in 1D ^{1}H NMR spectra. Spectral interference from buffer was minimized by use of $^{2}H_2O$. The use of $^{2}H_2O$ confers an additional advantage whereby leakage of saturation is minimized (Mayer and James, 2002; Begley *et al.*, 2013). In order to enhance the S/N, experiments were performed at 10 °C. Lowering the temperature improves the affinities of ligands that binds exothermically (Howard *et al.*, 2013), enhances the protein stability and compound solubility. Fragment cocktails that precipitated in the screening buffer were excluded. A typical STD spectrum from screening fragment cocktails is shown in **Figure 4.5**.

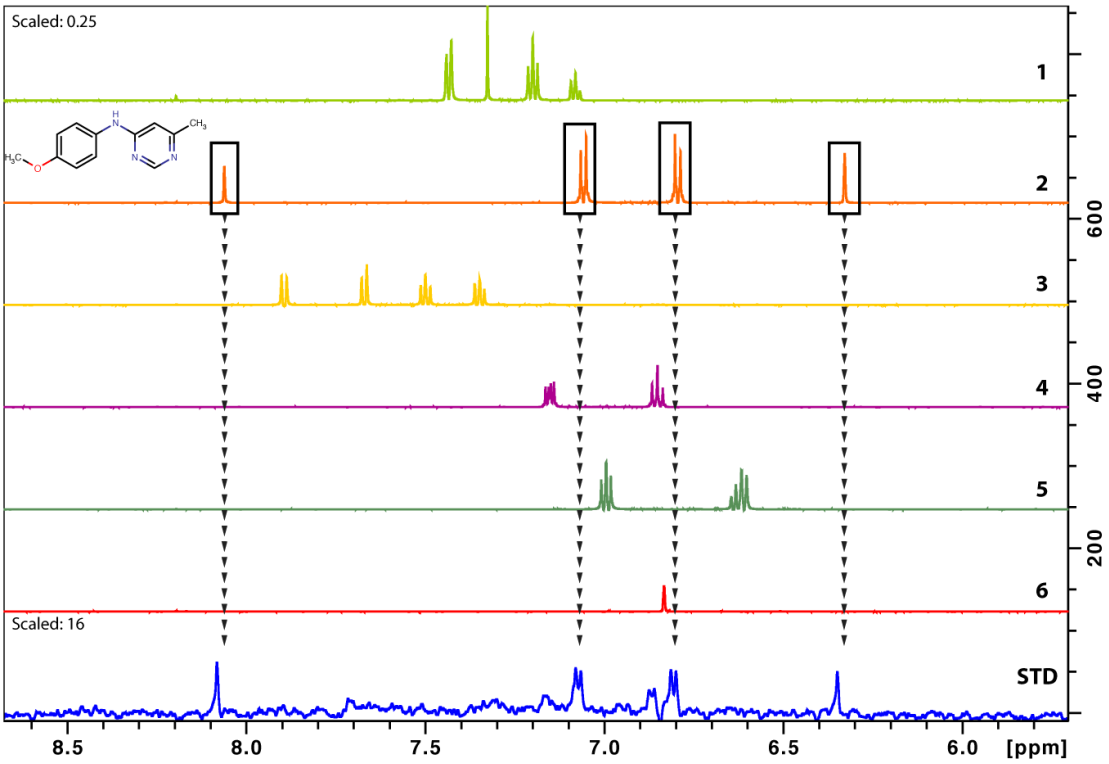


Figure 4.5. STD-NMR spectra of one of the rank 3 compounds. Spectra labelled 1-6 represent 1D ^{1}H NMR spectra of each individual compound present in the cocktail. In blue is the STD spectrum in the presence of protein. Arrows highlighting peaks indicating the peaks appearance in STD spectrum confirming that second compound binds to protein from a cocktail.

Assessment of STD experiments was based only on the aromatic region of the spectrum. In order to analyse the STD experiments and differentiate the quality of binding, a ranking system based on comparison of the STD signal intensity was applied: *rank 0*-no STD; *1*-possible hit, very week intensity; *2*-clear hit, moderate intensity and *3*-clear hit, strong intensity) (Doak *et al.*, 2013). Fragments with a positive STD signal from the cocktail screen were re-tested as pure ligands using the same technique. The overall STD hit rate was 17%: 56 compounds were *rank 3* (clear hit, strong intensity), 62 were *rank 2* (clear hit, moderate intensity), and 82 were *rank 1* (possible hit, weak intensity). The chemical structures of identified fragments are listed in *Appendix B*. Of the *rank 3* compounds, 26 (about half) were identified as "frequent hits" (Doak *et al.*, 2013) and were excluded from further consideration. Similarly, 24 of the *rank 2* hits were excluded.

The STD-NMR experiments were done primarily to identify fragments binding to SSB-Ct binding pocket of DnaG primase. As the protein used contained the RNA polymerase domain further experiments are required to distinguish those compounds that bind to the RPD domain.

4.2.4 Chemistry of hits

Those hits (from both libraries) showing interaction with RCD by two different biophysical methods display some structural similarities (**Figure 4.6**). Apart from having hydrophobic moieties such as phenyl rings and heterocyclic groups (commonly indole, diazole, benzodiazole and oxygen containing heterocycles etc.), some of them also include sulphonamide and carboxylic acidic motifs that could be mimicking the carboxylic acid moiety of Phe177 in SSB-Ct. Residues engaged in salt bridge interactions with the carboxy-terminus of SSB-Ct are conserved across binding partners, as shown in crystal structures and MD simulations described in *Chapter 3*.

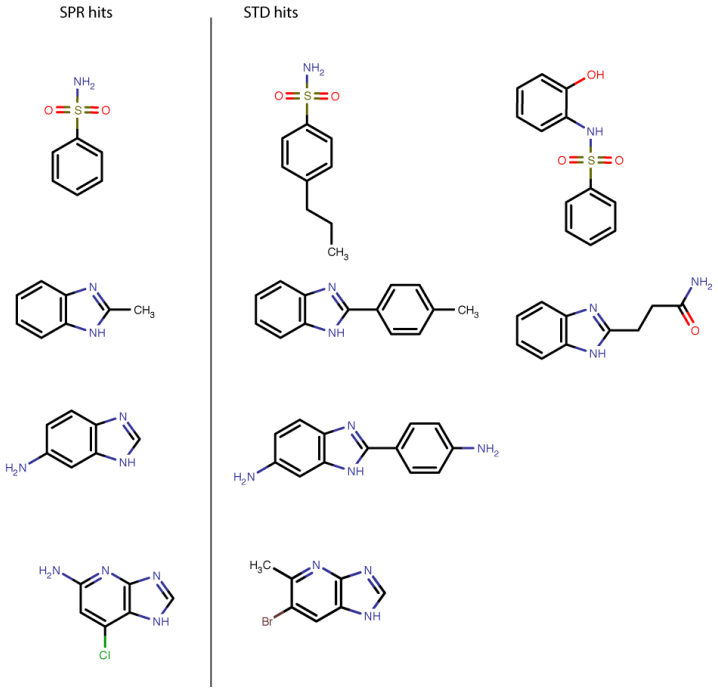


Figure 4.6. Comparison of some of the SPR and STD hits.

4.3 Conclusion

Initially an attempt to use crystallography for fragment screening against DnaGC was made. As no suitable crystals of DnaGC or RCD proteins were forthcoming, SPR and STD-NMR approaches were used. The Zenobia "first pass screen", containing 350 fragments, was screened in a SPR competition assay. Hits from the SPR experiments, and the MIPS fragment library were rescreened by STD-NMR. Fragment hits against the SSB-Ct binding pocket and RNA polymerase domain were identified.

Chapter 5

Structural and Biophysical Characterisation of Hits

5.1 Introduction

The next step in this project was to identify the binding pocket(s) of identified fragments on the surface of the target protein. Protein-based NMR methods can give information about ligand binding and dynamics of the complex and in some cases can give rise to NOEs which are transferred to distance restraints (Petros *et al.*, 2014).

Several methods have been developed for studying proteins and protein/protein interactions by NMR, the most common being $^{15}N/^{13}C$-HSQC. The availability of cryogenic probes and improved pulse sequences make it possible to obtain protein resonances at lower concentrations of protein (e.g. at 50 μM using SOFAST-HSQC) (Schanda *et al.*, 2005). High acquisition speeds allow for high-throughput experiments. Using TROSY-based experiments with various isotopically selectively labelled samples allows the method to be applied to larger proteins (> 100 kDa) (Riek *et al.*, 2000). With ^{15}N-labelled biomolecules, it provides a high-resolution 'fingerprint' of the protein. Each amino acid except proline is represented by a single backbone amide peak. An extra peak is observed for the indole-containing side chain of tryptophan and two extra peaks are observed for the side chains of asparagine, glutamine. In certain cases an additional peak is observed for side chains lysine and arginine (Kwan *et al.*, 2011). A 2D-HSQC spectrum of ^{15}N-enriched protein with ligand allows ligand-binding locations to be determined. Where binding occurs, changes in chemical shifts are observed. Chemical shift perturbation (CSP) mapping provides crude structural information. Chemical shifts are precise, reproducible and are very sensitive to their local environments. Therefore, small changes in structure and/or dynamics are readily detectable (Williamson, 2013; Pellecchia *et al.*, 2002). An important caveat is that CSPs may be due to allosteric effects. Thus, CSP may not exclusively show the exact binding site. In such cases, comparison of spectra of two similar ligands is recommended (Pellecchia, 2005).

In cases where crystal structures of bound fragments are not available, molecular docking may provide insights into how the ligand is oriented in the pocket. The main disadvantage of "unrestrained" docking calculations is that discriminating between native and non-native orientations is difficult (Wang *et al.*, 2007). The presence of restraints (e.g. CSP, NOE-restrains) significantly improves the accuracy of molecular docking. CSPs can be used to generate docking restraints. However, CPS data is most commonly used as a final filter of poses generated by docking (Stark and Robert

Powers, 2008) Medek and co-workers have shown that similar ligands with similar binding poses generate similar CSPs (Medek *et al.*, 2000).

^{19}F-NMR can also provide valuable data for the study of protein-ligand interactions. The high sensitivity and specificity of the method makes it a valuable tool for structural biology and drug design (Chen *et al.*, 2013). ^{19}F has ½ spin without a quardrupole moment, thereby significantly simplifying the interpretation of spectra. It's short longitudinal relaxation rate (R_1) makes spectral acquisition relatively fast. It has a natural abundance of 100% and a high gyromagnetic ratio, resulting in strong dipolar coupling (e.g. ^{19}F-^{19}F and ^{1}H-^{19}F NOEs) that can be used as distance restraints in structural calculations. Fluorine-19 displays a wide range of chemical shifts (-300 to 400 ppm). It allows precise detection of changes in chemical environments. Its NMR signal is free of background signals from proteins and/or buffers. The acquisition of ^{19}F spectra does not require the use of deuterated buffers and is therefore cost-effective. While fluorine has a larger van der Waals radius compared to hydrogen (1.35 Å *vs* 1.10 Å), studies have shown that the additional steric demand of fluorine in comparison to hydrogen is small (Duschinsky *et al.*, 1957). Incorporation of fluorinated amino acids into proteins (in aliphatic or aromatic side chains) does not usually cause structural and functional changes (Didenko *et al.*, 2013; Vulpetti *et al.*, 2009; Dalvit *et al.*, 2004). The replacement of hydrogen with fluorine in drugs has many advantages. Typically, its influence on the metabolic profile of compounds is to block oxidation (Hagmann, 2008; Purser *et al.*, 2008). These properties of ^{19}F nuclei render it a useful screening tool. It is noteworthy that the S/N ratio of 1D ^{19}F spectra can be improved by use of cryogenic probes and by increase of the field strength of the spectrometer.

5.1.2 Aims

The aim is to validate the first generation hits using protein-based NMR techniques, to identify binding sites on DnaGC, the elucidation of fragment binding-modes and biophysical characterisation of binding. Furthermore, hit-to-lead optimisation using *in silico* approaches were performed.

5.2 Results and Discussion

5.2.1 ^{15}N-HSQC Buffer Optimisation of DnaGC

The structure of DnaGC was determined by crystallography and NMR (Oakley *et al.*, 2005; Su *et al.*, 2006). There are two NMR experimental conditions described for DnaGC (1) 10 mM phosphate buffer (pH 6.0), 100 mM NaCl, 1 mM DTT and 0.1% (w/v) sodium azide (Su *et al.*, 2006), and (2) 5 mM potassium phosphate (pH 7.4), 5 mM NaCl, 2 mM DTT (Naue *et al.*, 2013). The latter condition was used to identify the SSB-Ct binding pocket, although certain residues, Lys447, Arg448, Thr449, Asn511, near the SSB-Ct binding site, and Asn565 in the hairpin turn near the C-terminus were not detected in ^{15}N-HSQC spectrum of the apo-protein. Resonances due to those residues appeared in the spectrum only in the presence of an excess of SSB-Ct peptide (Naue *et al.*, 2013). Therefore, condition 1 was chosen. To avoid possible pH changes caused by addition of some of the ionisable ligands, condition 1 was slightly modified by replacing the phosphate by 50 mM MES, which is recommended for pH range between 5.5 to 6.5 (Kelly *et al.*, 2002). The 2D ^{15}N-HSQC spectrum of DnaGC in the optimised buffer is shown in **Figure 5.2**.

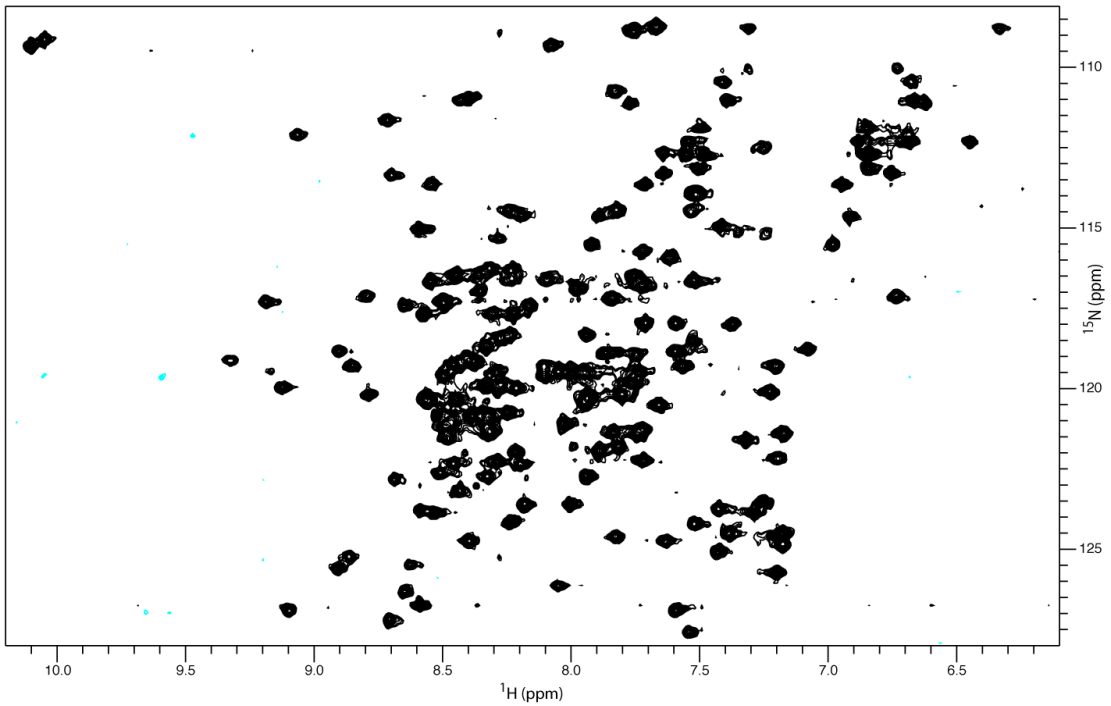


Figure 5.2. ^{15}N-HSQC spectrum of DnaGC. Spectrum was recorded at 298 K using 100 µM protein.

All backbone amide peaks in ^{15}N-HSQC were assigned and in most cases showed some small shifts relative to the spectrum reported by Su and co-workers (Su *et al.*, 2006).

5.2.2 ^{15}N-HSQC Assignment Propagation of DnaGC and First Generation of Hits

The influence of organic solvents in shifting peaks in NMR spectra is well known (Liepinsh and Otting, 1997). In order to minimise the possibility of false positive results, and to exclude the effects of DMSO on CSPs, DMSO was mixed with ^{15}N-DnaGC and a HSQC spectrum was acquired. DMSO concentrations of up to 2% (v/v) did not shift any peaks. Rank 2 and 3 compounds (68 fragments; see Chapter 4) that were not classified as "frequent hits" were assessed for binding using ^{15}N-HQSC. Spectra were recorded for the apo-protein (100 µM) and after addition compounds (3.3 mM). Where binding occurs, peak shifts and/or intensity changes are expected (**Figure 5.3,** *Appendix C*). Weak protein-fragment interactions often have high dissociation rate constants (k_{off}), resulting in averaged chemical shifts from both ligand-free and ligand-bound states (i.e. separation of NMR signals for free and bound forms does not occur), which is typical for fast-exchange systems (Dalvit and Stockman, 2002).

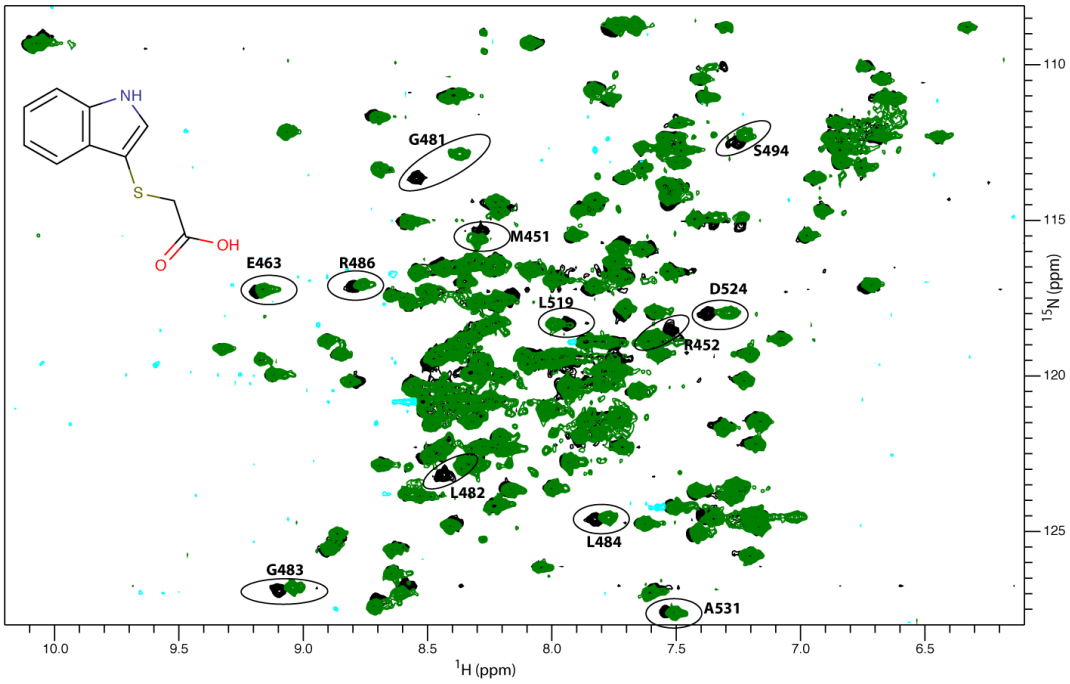
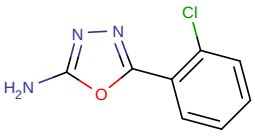
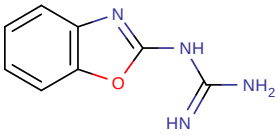
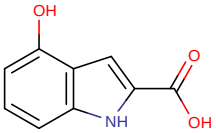
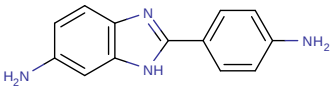
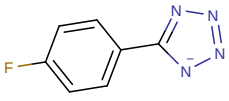
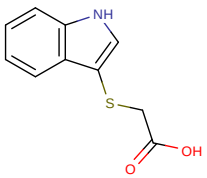
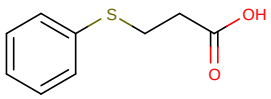
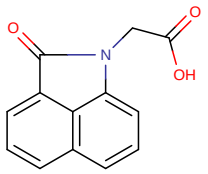
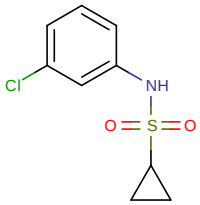
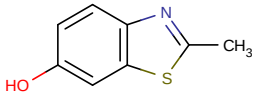
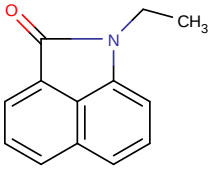
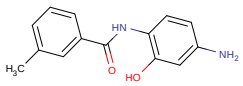


Figure 5.3. ^{15}N-HSQC spectrum of apo ^{15}N-DnaGC (black) and upon addition of 3.3 mM L1H10 fragment (green).

Therefore, chemical shift changes were assigned using the nearest-neighbour approach. By mapping the CSP data to the protein surface, the binding locations of the fragments were identified. The STD hits that were confirmed by ^{15}N-HSQC NMR measurements were ranked according to CSP magnitude (section 2.2.4.5) in order to identify the most potent fragment. Typically, perturbations were considered significant if $\Delta\delta > 0.02$ ppm for at least two peaks in the spectrum using **Eq. 2.2** (Hajduk *et al.*, 2000). 12 fragments were assigned as rank 3 and 2 by HSQC (**Table 5.2**), and 56 fragments gave no perturbations.

Table 5.2. The list of HSQC hits. Compounds were ranked according to CSP magnitude. The number of rotatable bonds and cLogP was estimated using Instant JChem 6.1.5, 2014, ChemAxon (http://www.chemaxon.com).

Name	Structure	HSQC rank	Number of rotatable bonds	cLogP
L3F10		3	1	1.44
L3H1		3	1	0.98
L4B4		3	1	1.35
C2G4		3	1	1.63
C4C4		3	1	1.4
L1H10		3	3	1.87
L1A11		3	4	2.4

Name	Structure	HSQC rank	Number of rotatable bonds	cLogP
L1G8		3	2	1.42
L6D1		3	2	1.72
L1F5		2	0	1.93
L1C6		2	1	2.30
C9C12		2	2	2.45

Comparing the CSP data for fragments confirmed by ^{15}N-HSQC NMR shows similar patterns of chemical shifts (albeit with different magnitudes), confirming that all fragments bind to the same pocket. Residues Arg452 Arg452 (^{15}N-118.54 ppm, ^{1}H-7.53 ppm), Lys518 (^{15}N-119.2 ppm, ^{1}H-7.8 ppm), Thr449 (^{15}N-114.79 ppm, ^{1}H-7.59 ppm) and Gly481 (^{15}N-113.9 ppm, ^{1}H-8.6 ppm) were most susceptible to perturbation upon addition of fragments (**Figure 5.4,** *Appendix C*). Number of rotatable bonds does not include bonds with hydrogen atoms.

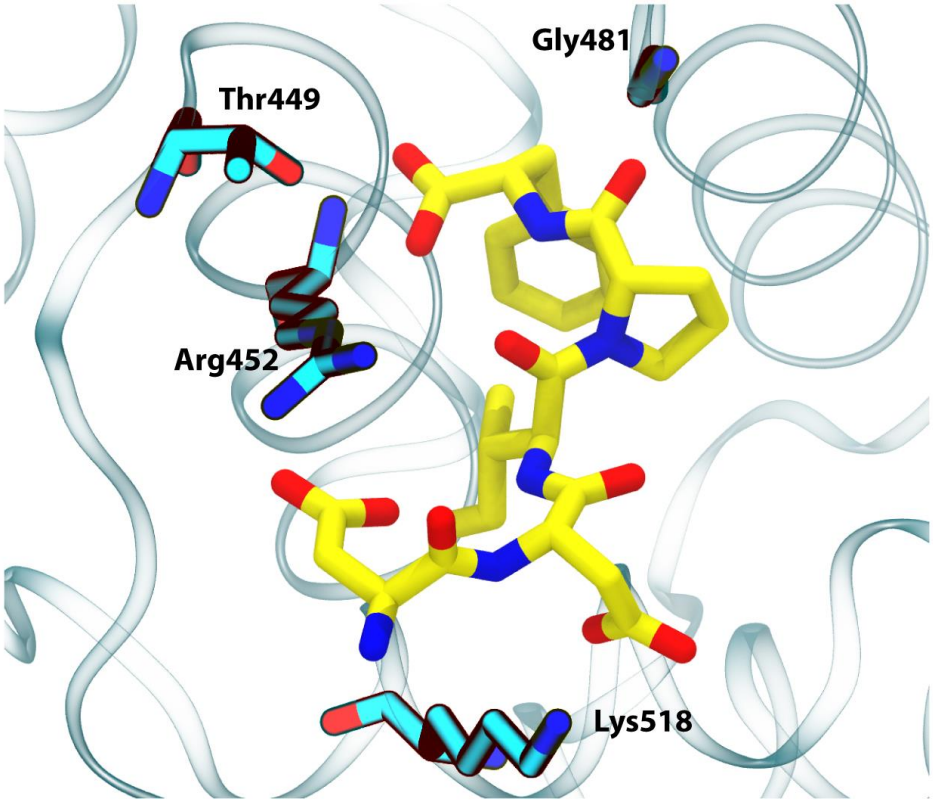


Figure 5.4. SSB-Ct binding pocket of DnaGC. Representation of residues most susceptible to CSP upon addition of fragments.

These residues are also vital for binding SSB-Ct peptide. Four fragments (*L1G8* (2-[2-oxobenzo[cd]indole-1(2H)-yl]acetic acid), *C4C4* (5-(4-fluorophenyl)-1H-1, 2, 3, 4-tetrazole), *L1H10* ((1H-indole-3-ylsulfanyl)acetic acid) and *L1A11* (3-(phenylsulfanyl)propanic acid)) (**Figure 5.5**) have been chosen as the starting point for further optimisation, based on strong CSP changes (*Appendix C*). In spite the fact that HSQC experiments identify the binding site, they do not provide the binding orientation of fragments. The binding site location can nevertheless be used for CSP-guided molecular docking to predict the binding orientation of fragments.

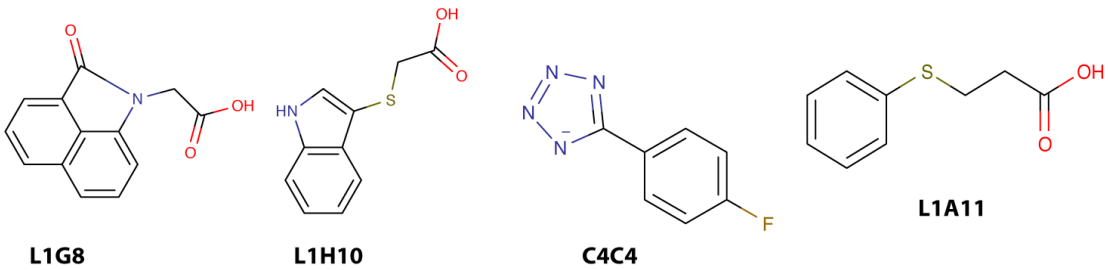


Figure 5.5. The four best fragments estimated binding affinities identified from primary screening. Binding affinities were estimated by ^{15}N-HSQC titration.

5.2.3 Biophysical Characterisation of Confirmed Hits

Binding affinities of fragments were estimated by HSQC titration experiments, without exceeding the 2% (v/v) DMSO level used previously. Peak movement were monitored throughout the titration and, chemical shift changes were measured for each single concentration point (**Figure 5.6**).

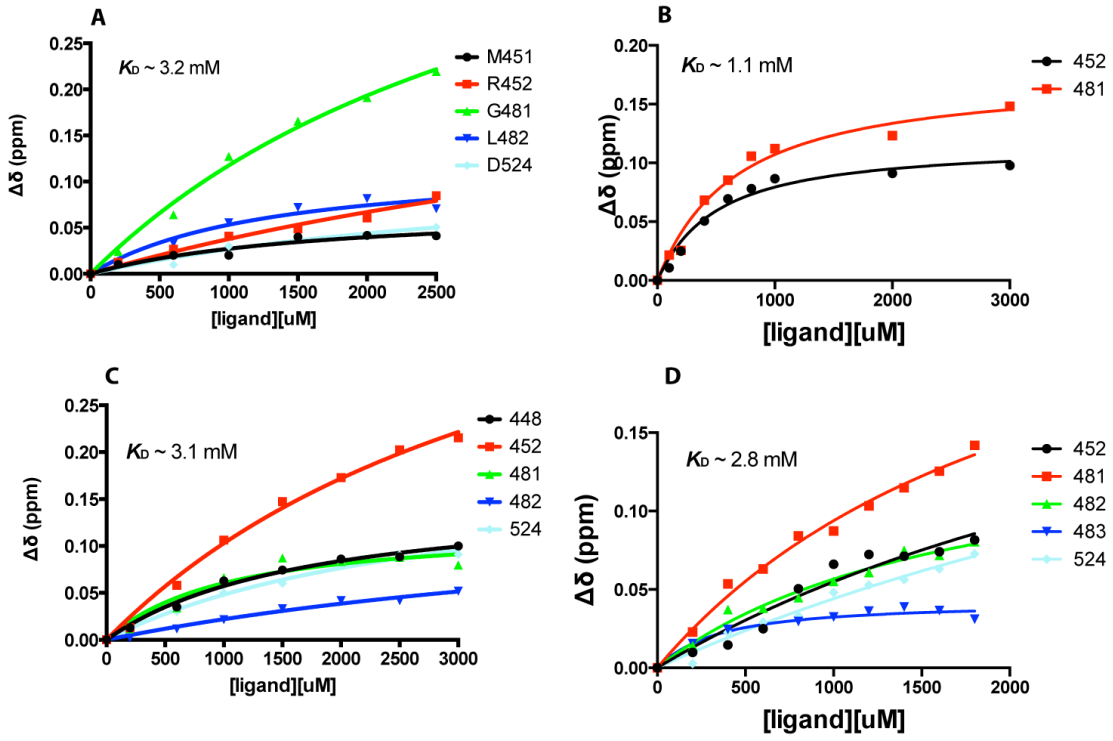


Figure 5.6. Binding affinities measured by HSQC titration of (A) C4C4, (B) L1G8, (C) L1H10 and (D) L1A11. Different curves corresponds to different perturbed residues upon addition of fragment.

All four fragments (C4C4, L1H10, L1A11) showed similar affinities to DnaGC. The highest affinity was shown for the L1G8 fragment, possibly as a result of occupying the shallow binding pocket with its hydrophobic skeleton (discussed in 5.2.4). It is often difficult to measure the binding affinity of weak binding ligands caused by protein and/or ligand solubility and as a consequence K_D is underestimated (Fielding, 2007).

5.2.4 CSP-Guided Molecular Docking

In the absence of experimental data, computational approaches can be used to predict the structures of protein/ligand complexes. Using molecular docking calculations, it is possible to predict the orientations of small molecules in binding pockets. Docking algorithms generate possible orientations of ligands, which when complimented with

experimental results from protein-based NMR methods, may assist in the identification of the most likely ligand orientations within a binding pocket.

The crystal structure of the DnaGC/SSB-Ct complex was used to dock the fragments. For this purpose, Autodock Vina 1.1.2 was used (Trott and Olson, 2010). It is computationally efficient and gives relatively low standard errors of free energy of binding (compared with theoretical and experimental free energy of binding) of 2.85 kcal/mol (Trott and Olson, 2010). It uses combined knowledge-based potentials and empirical scoring functions. It has also shown higher success rates and computational efficiency compared to other docking programs (Krüger *et al.*, 2012).

The crystal structure of DnaGC/SSB-Ct complex revealed that there is no major conformational changes upon binding of SSB-Ct peptide to DnaGC (Tak Lo, 2012), therefore in all docking calculations, the protein remained static and flexibility was afforded to ligand molecules. The docking grid covered the SSB-Ct binding pocket, with dimensions 16 Å x 16 Å x 14 Å, centred on the binding location of SSB-Ct residue Phe177. The 4 fragments selected based on the HSQC experiments above were docked, and the first three low energy conformers are shown in **Figure 5.7**. Only first pose of the docked fragments was considered for detailed description of interactions with the binding pocket of DnaGC (Dr. Haibo Yu personal communication). The binding free energy is shown in **Table 5.3**.

Table 5.3. The binding free energy of docked poses of first generation of fragments generated by Autodock Vina. Binding free energies are in kcal/mol.

	Pose 1	**Pose 2**	**Pose 3**	**Pose 4**	**Pose 5**	**Pose 6**	**Pose 7**	**Pose 8**	**Pose 9**	**Pose 10**
C4C4	-5.1	-5.1	-5.1	-5.0	-4.9	-4.9	-4.7	-4.7	-4.7	-4.7
L1G8	-5.7	-5.5	-5.5	-5.4	-5.3	-5.3	-5.2	-5.2	-5.2	-5.2
L1H10	-5.4	-5.1	-5.0	-5.0	-4.9	-4.9	-4.8	-4.8	-4.6	-4.6
L1A11	-4.7	-4.7	-4.6	-4.6	-4.6	-4.5	-4.5	-4.5	-4.5	-4.5

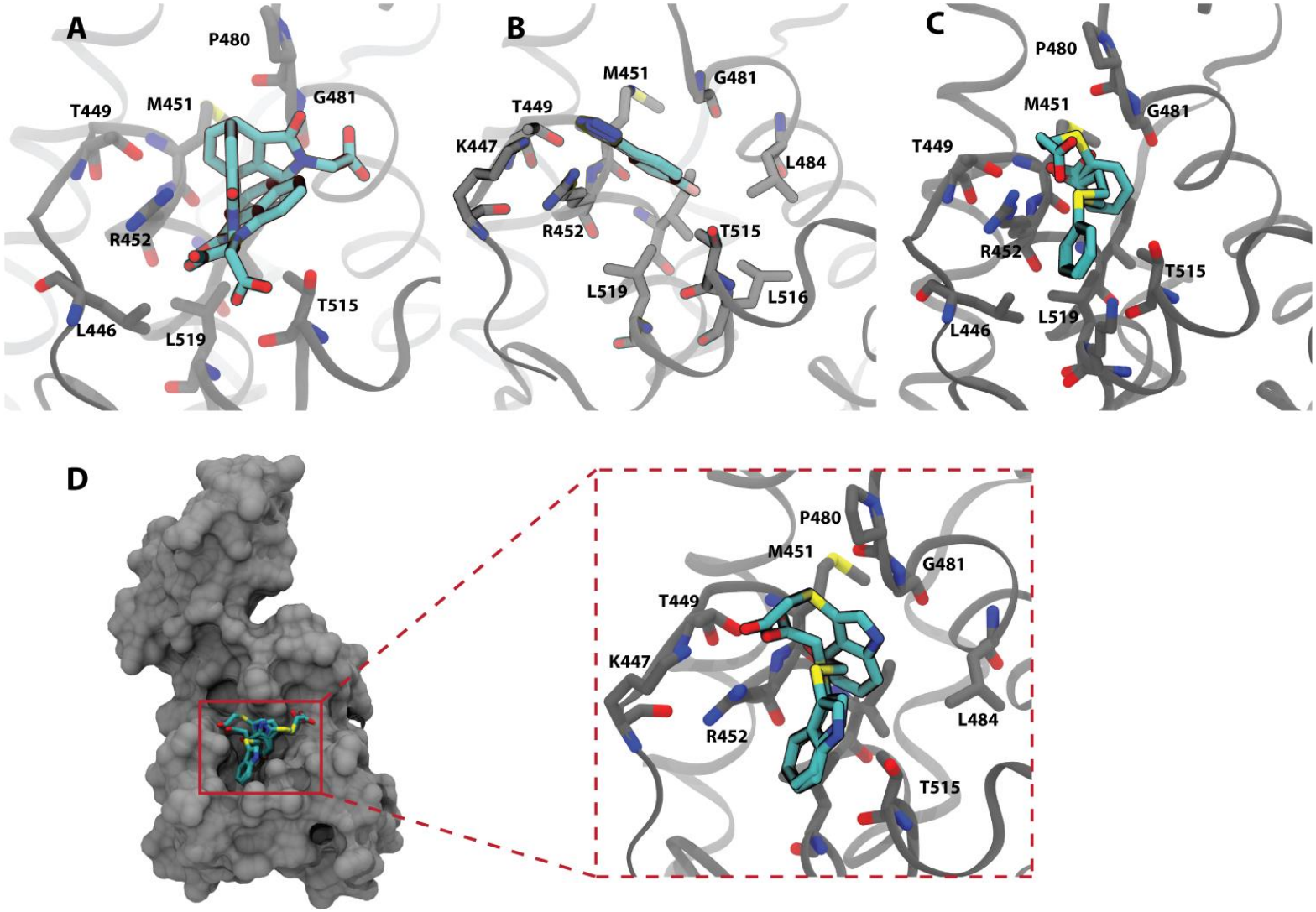


Figure 5.7. Predicted orientations of the best four first generation of hits (L1G8 (A), C4C4 (B), L1A11 (C) and L1H10 (D)). The 3 lowest energy poses shown. Figure was generated using VMD (Humphrey et al., 1996).

The MD simulations described in *Chapter 3* and the crystal structure of DnaGC/SSB-Ct complex indicate that Arg452 is extensively involved in hydrogen bonding/electrostatic interactions with carboxylic acid group of Phe177. CSP-guided molecular docking suggests that all 4 compounds interact with Arg452 (^{15}N-118.54 ppm, ^{1}H-7.53 ppm). Detailed examination of a predicted orientation of L1G8 fragment is shown in **Figure 5.8**. The shallow hydrophobic pocket is occupied by the tricyclic skeleton (**Figure 5.8A**).

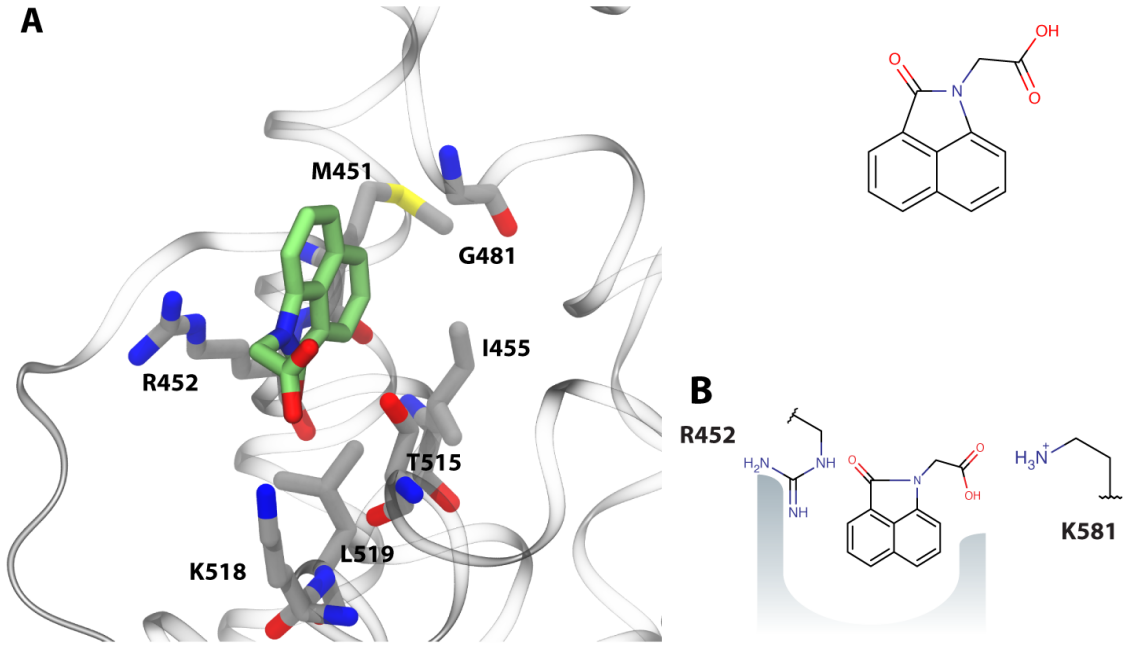


Figure 5.8. NMR-guided docking of L1G8 fragment. (A) Lowest energy docked binding pose of L1G8. Schematic representation of binding pocket with charged interactions between protein and fragment (B).

While the carbocyclic skeleton forms hydrophobic contacts with the pocket, the carboxylic group forms a salt bridge with the positively charged side chain of Lys518 (^{15}N-119.2 ppm, ^{1}H-7.8 ppm). The carbonyl oxygen of this fragment forms a hydrogen bond with Arg452. Interestingly, one of the other fragment hits, L1C6, has the same chemical skeleton as L1G8, but with the carboxylic acid group replaced by a methyl group. Spectral overlay of two compounds in complex with DnaGC, ^{15}N-HSQC spectra indicates that fewer peaks have shifted compared with L1G8, in agreement with molecular docking. Additionally CSP-magnitudes are smaller compare to L1G8 (**Figure 5.9,** *Appendix C*). The data suggests that charged interactions of the L1G8 carboxylic acid motif contribute significantly to the interaction with DnaGC.

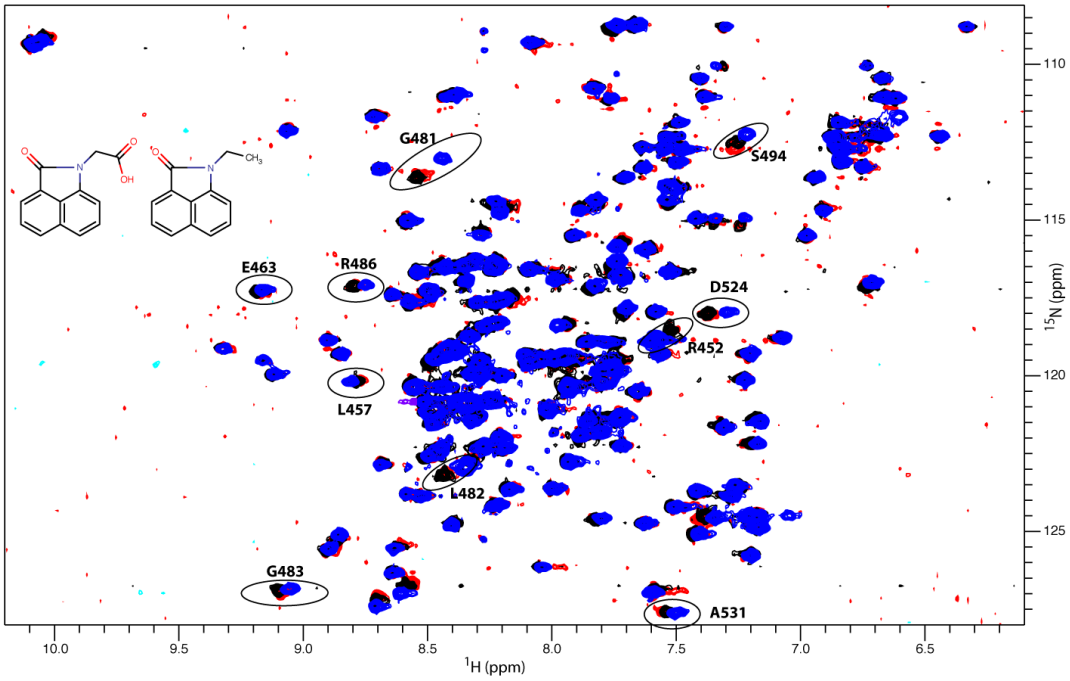


Figure 5.9. Comparison of CSP induced by L1C6 (red) and L1G8 (blue) compounds with chemical structures of two fragments. The apo-protein spectrum is in black.

CSP-guided molecular docking of C4C4 compound with DnaGC is shown in **Figure 5.10**.

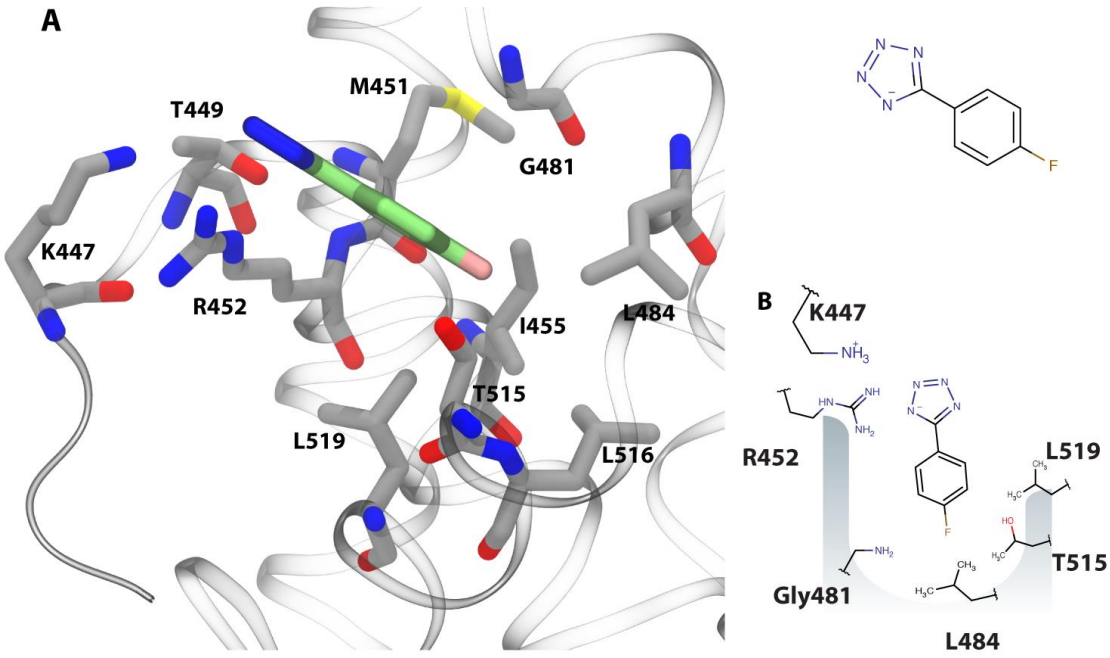


Figure 5.10. C4C4 docked into DnaGC. The lowest energy docked pose of the fragment in the DnaGC pocket is shown (A), and schematic representation of residues interacting with fragment (B).

In spite that fact that L1G8 has better binding affinity, C4C4 is predicted to make more contacts with the binding pocket that are potentially exploitable in lead optimisation

(*Appendix C*). Compound C4C4 contains a tetrazole group, a bioisosteric alternative to the carboxylic acid group (Malik *et al.*, 2013). It has a low p*K*a (~5) and is good hydrogen bond acceptor. Modelling studies suggest that replacement of carboxylic acids with tetrazole increases the hydrogen bond strength (Allen *et al.*, 2012). The tetrazole ion forms electrostatic and hydrogen bond interactions with the side chains of Lys447, Thr449 and Arg452. The phenyl ring appears likely to mimic the last phenylalanine residue (Phe177) in the SSB-Ct peptide and makes hydrophobic interactions with residues Pro480, Gly481, Thr515 and Leu519. The presence of a fluorine atom is expected to increase the acidity of the tetrazole due to its electron-withdrawing nature (Dalvit *et al.*, 2014). The fluorine atom is in contact with the amide group of Gly481 and the methyl groups of Leu448, Leu455, Thr515, Leu516 and Leu519 (**Figure 5.10**). These residues are forming the edge of the peptide binding pocket.

Fragment L1A11 has a slightly different binding pose compared to other three fragments. The carboxylic motif is directed towards the edge of the pocket and forms hydrogen bond interactions with amide group of Gly481. In the docked model of L1A11, the sulfur atom simultaneously interacts with the guanidine group (3.6 Å) and a backbone amide proton (3.2 Å) of Arg452 (**Figure 5.11**).

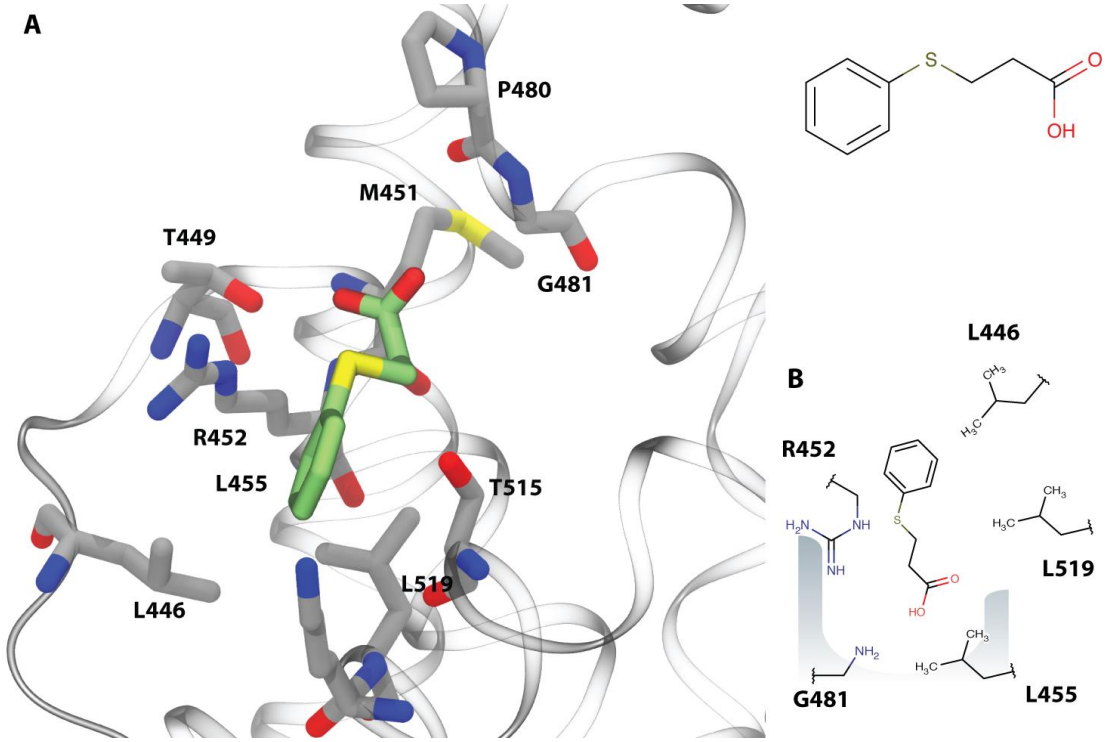


Figure 5.11. Visualization of binding of L1A11 fragment. Lowest energy orientation of fragment, in the DnaGC pocket (A). Schematic representation of residues interacting with fragment (B).

In case of the L1H10 fragment the aromatic ring is surrounded by Leu446, Thr515 and Leu519 forming various hydrophobic σ-π contacts. The indole ring is buried in the hydrophobic pocket formed by methyl groups of Ile455, Gly481, Leu484 and Leu519. Meanwhile Gly481 interacts with the nitrogen atom of indole (**Figure 5.12**). The carboxylic group interacts with Lys447 and Arg452.

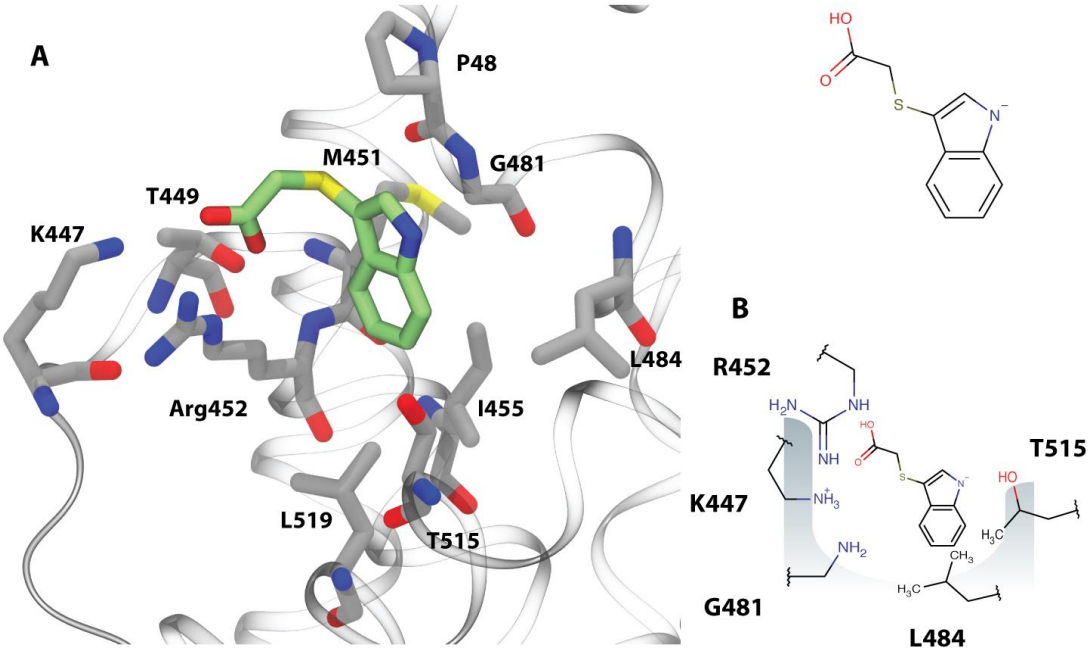


Figure 5.12. Visualization of binding of L1H10 fragment. Lowest energy dockedorientation of fragment in the DnaGC pocket (A). Schematic representation of residues interacting with fragment (B).

Fragments LA11 and L1H10 both contain carboxylic acid and thioether groups. Thioethers have a hydrophobic character, but sulphur is also a very weak hydrogen bond acceptor. The importance of anionic motifs was been demonstrated in the cases of L1G8 and C4C4 fragments. A negative charge on the ligand therefore appears necessary for targeting the SSB-Ct/DnaGC interactions.

5.2.5 In Silico Hit to Lead Optimisation

2D NMR confirmed that most of the fragment hits were binding to the SSB-Ct binding pocket of DnaGC. Autodock Vina software (Trott and Olson, 2010) was used for a high-throughput analogue screen. The hits confirmed by ^{15}N-HSQC experiment were used as starting point for the identification of structurally similar compounds to all four hits described above in the ZINC database (Irwin and Shoichet, 2005). About 7700 compounds were selected for docking. The Y-shape geometry of the pocket gives potential vectors to grow molecules (**Figure 5.13**).

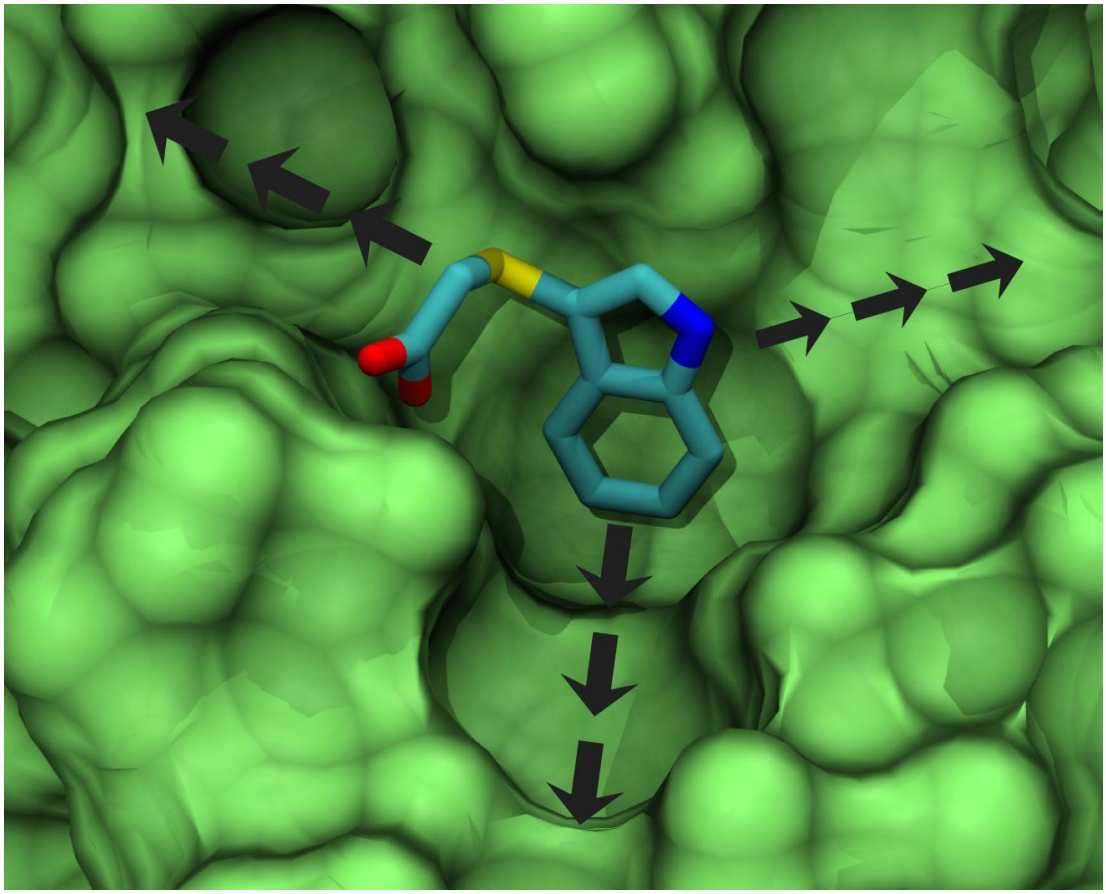


Figure 5.13. The pocket geometry of DnaGC with predicted orientation of L1H10 fragment. The black arrows indicating possible vectors for compound optimisation.

Docking identified about 30 commercially available compounds with promising binding poses and favourable binding energies. These were purchased and the binding to DnaGC assessed using STD and 2D NMR experiments (listed in *Appendix D*). Some of the compounds were not soluble in DMSO and were therefore excluded from screening. Some of the remaining analogues had low solubility limits, therefore the highest concentrations were either 200, 500 μM or 1 mM, depending on the compound.

The docked C4C4 fragment has good vectors for fragment growth. With the knowledge that the tetrazole moiety might have improved membrane-crossing properties compared with carboxylates, owing to a higher p*Ka*, the C4C4 fragment was selected for further development. The *in silico* analogue screen identified 10 tetrazole analogues (listed in *Appendix D*) with favourable binding poses relative to SSB-Ct peptide. From STD and 2D NMR experiments most of the compounds obtained did not bind to DnaGC, although 5-[2-fluoro-6-(4-fluorphenoxy)phenyl]tetrazol-1-ide (ZINC72447025) showed an STD signal, and altered CSP in 2D NMR experiments. The NMR spectrum of this

compound superimposed over the apo-protein spectrum is shown in **Figure 5.14**. The predicted binding free energy of optimised compounds is shown in **Table 5.4**.

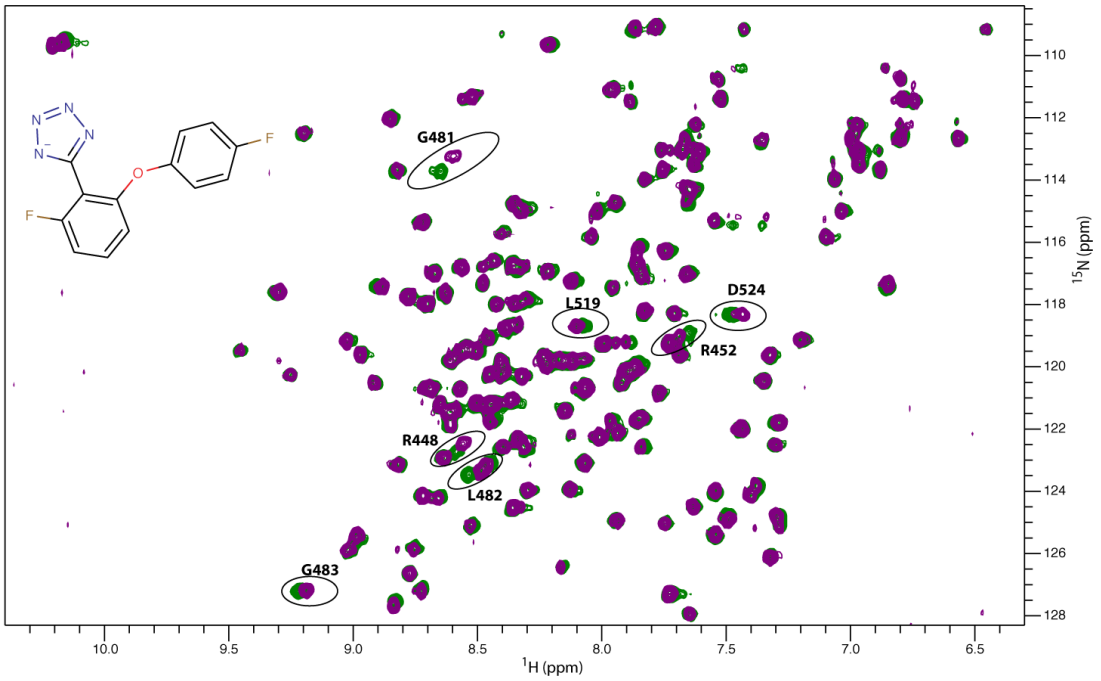


Figure 5.14. ^{15}N-HSQC spectra of apo-protein overlayed with ZINC72447025 compound spectrum. Green represents the apo-spectrum, magenta in the presence of 1 mM ZINC72447025 compound.

Table 5.4. The binding free energy (kcal/mol) of docked poses of first generation of fragments.

	Pose 1	**Pose 2**	**Pose 3**	**Pose 4**	**Pose 5**	**Pose 6**	**Pose 7**	**Pose 8**	**Pose 9**	**Pose 10**
C4C4	-5.1	-5.1	-5.1	-5.0	-4.9	-4.9	-4.7	-4.7	-4.7	-4.7
ZINC72447025	-7.0	-7.0	-6.8	-6.8	-6.6	-6.1	-6.1	-6.1	-6.0	-6.0
CDS001350	-6.7	-6.7	-6.6	-6.3	-6.3	-6.2	-6.2	-6.2	-6.2	-6.2
CHEMBR7179770	-6.3	-6.3	-6.3	-6.3	-6.2	-6.1	-6.1	-6.1	-6.0	-6.0

Mapping the CSP on protein surface (*Appendix C*), and using that data, predicted binding pose of ZINC72447025 compound was calculated (**Figure 5.15**).

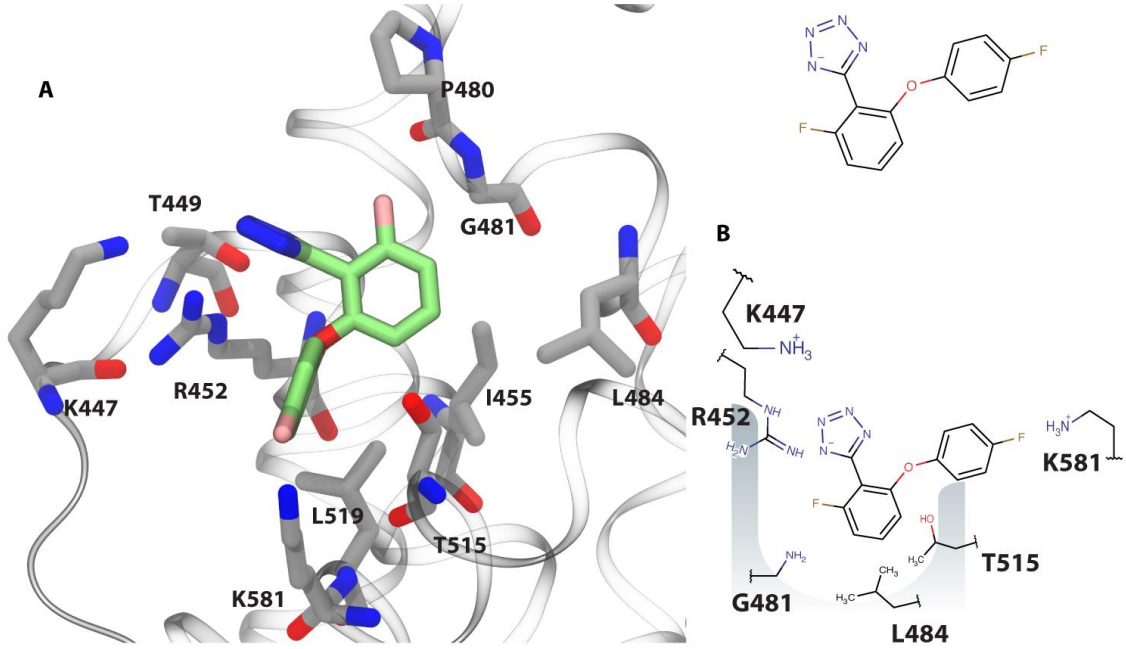


Figure 5.15. Visualisation of binding of ZINC72447025 compound. The lowest energy binding poses of ZINC72447025 (**A**). Schematic representation of residues involved in interaction with compound (**B**).

The observed NMR peak shifts were consistent with the docked orientation (*Appendix C*). The negatively charged tetrazole is predicted to form favourable electrostatic and hydrogen bond interactions with the side chains of Lys447 and Arg452. The methyl groups of Leu446, Met451, Leu455, Leu484 and Thr515, Leu519 form a hydrophobic pocket accommodating two aromatic rings of the compound. One of the fluorine atoms is about 3 Å from the amide group of Gly481, while the second one at the *para*-position of the phenoxy group points out of the pocket toward Lys518. These fluorine atoms could be involved in hydrogen bond formation with Gly481 and Lys518 (**Figure 5.15**). In order to further test the docked orientation of ZINC72447025 compound 1D ^{19}F NMR has been carried out in a 20-fold excess of compound. Fluorine signals were broadened (by 0.4 Hz and 1.8 Hz respectively) and shifted slightly downfield in 1D ^{19}F spectra indicating the binding shown in **Figure 5.16**.

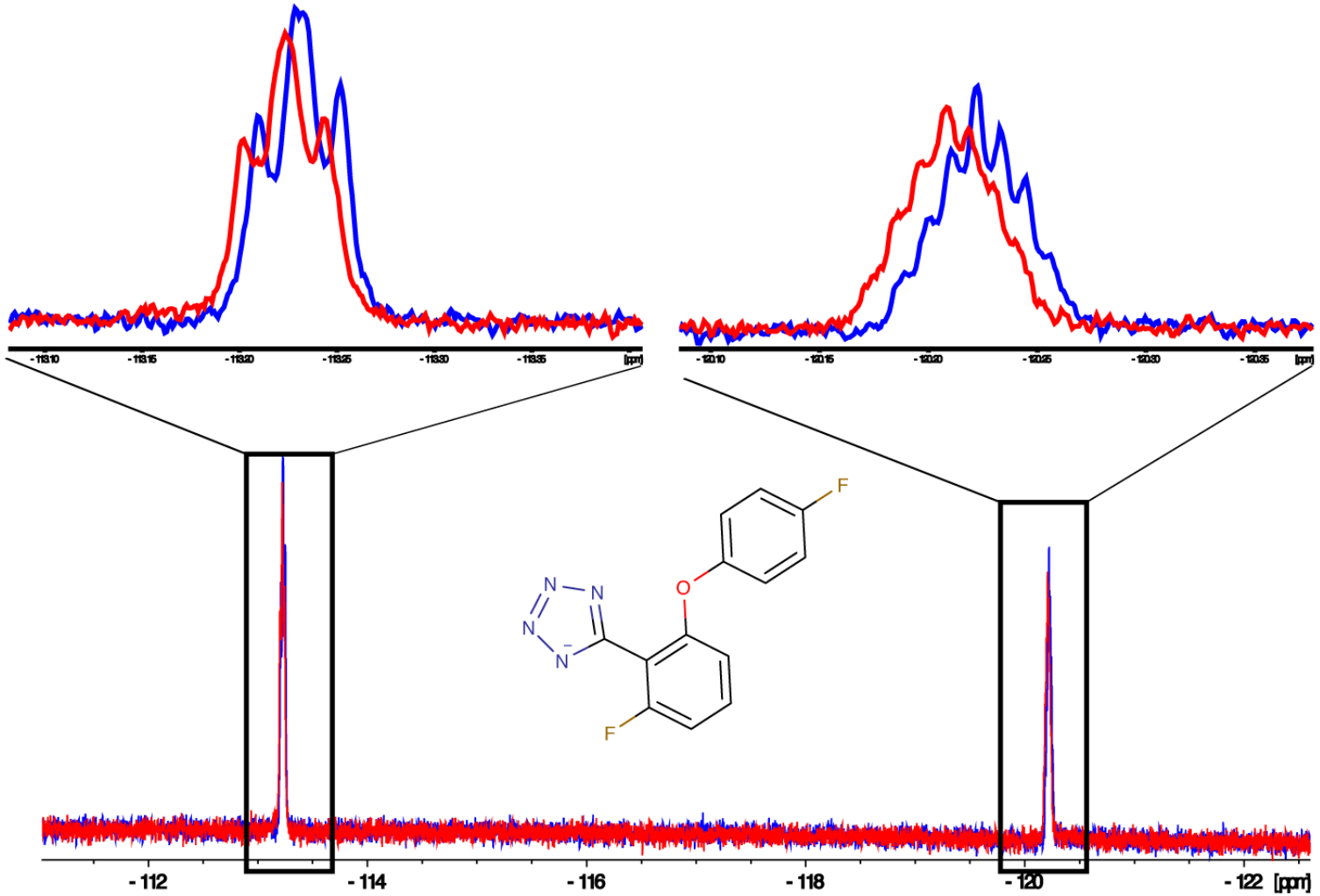


Figure 5.16. 1D ^{19}F spectra of ZINC72447025 compound at 1mM in presence (red trace) and absence (blue trace) of 50 µM DnaGC.

The model suggested that one fluorine atom is in close proximity (3 Å) to backbone amide hydrogen of Gly481. The most common NMR approach to gain distance restraints are NOEs. Inter-molecular NOEs are usually observed between complexes with tight binding (K_D < 10 uM) in the slow exchange regime (off-rate of the exchange process should be smaller than the NOE defined by the cross-relaxation rate constant) (Neuhaus and Williamson, 1984; Vögeli, 2014). To obtain NOE-based distance restraints, a 3D ^{15}N-NOESY-HSQC spectrum was acquired. Where the ligand is located less than 6 Å from backbone protons, ligand-to-protein NOEs will be observed. However, in a weak protein-ligand interactions, because of the fast exchange between free and bound states of protein-ligand complexes, the population of bound states was small, therefore no protein-ligand NOEs were observed. Although the side chain region of the spectrum had a new NOE peak, it was not assigned and unfortunately could not be used.

^{15}N-HSQC experiments were used to measure the binding affinity of the first generation of optimised hits described in section *2.2.4.7*. The identified ZINC72447025 tetrazole derivative was shown to bind to DnaGC with about 3-fold improved affinity compare to starting C4C4 fragment, as measured by HSQC titration (1.3 mM) (*Appendix E*). Binding affinities are measured by averaging among few resolved resonances in HSQC spectra and this could lead to noticeable inaccuracies (some of the peaks in HSQC spectra may be in different exchange rates). Additionally, limited ligand solubility, aggregation at high concentrations often does not allow to rich the saturation during the titration, thus resulting in underestimated K_Ds. Therefore, binding affinities for such weak interactions as estimated by HSQC titration should be treated with caution (Fielding, 2007; Lian, 2013).

Searching the ZINC database (Irwin and Shoichet, 2005) for ligands structurally similar to ZINC72447025 indicated that the only commercially available analogue of was CDS001350. CDS001350 is missing the halogen atoms and has *para*-phenoxy group (**Figure 5.17**). The binding of CDS001350 to DnaGC was assessed by STD NMR followed by ^{15}N-HSQC **Figure 5.17A**. The overlaid HSQCs of ZINC72447025 compound with CDS001350 shown in **Figure 5.17B**.

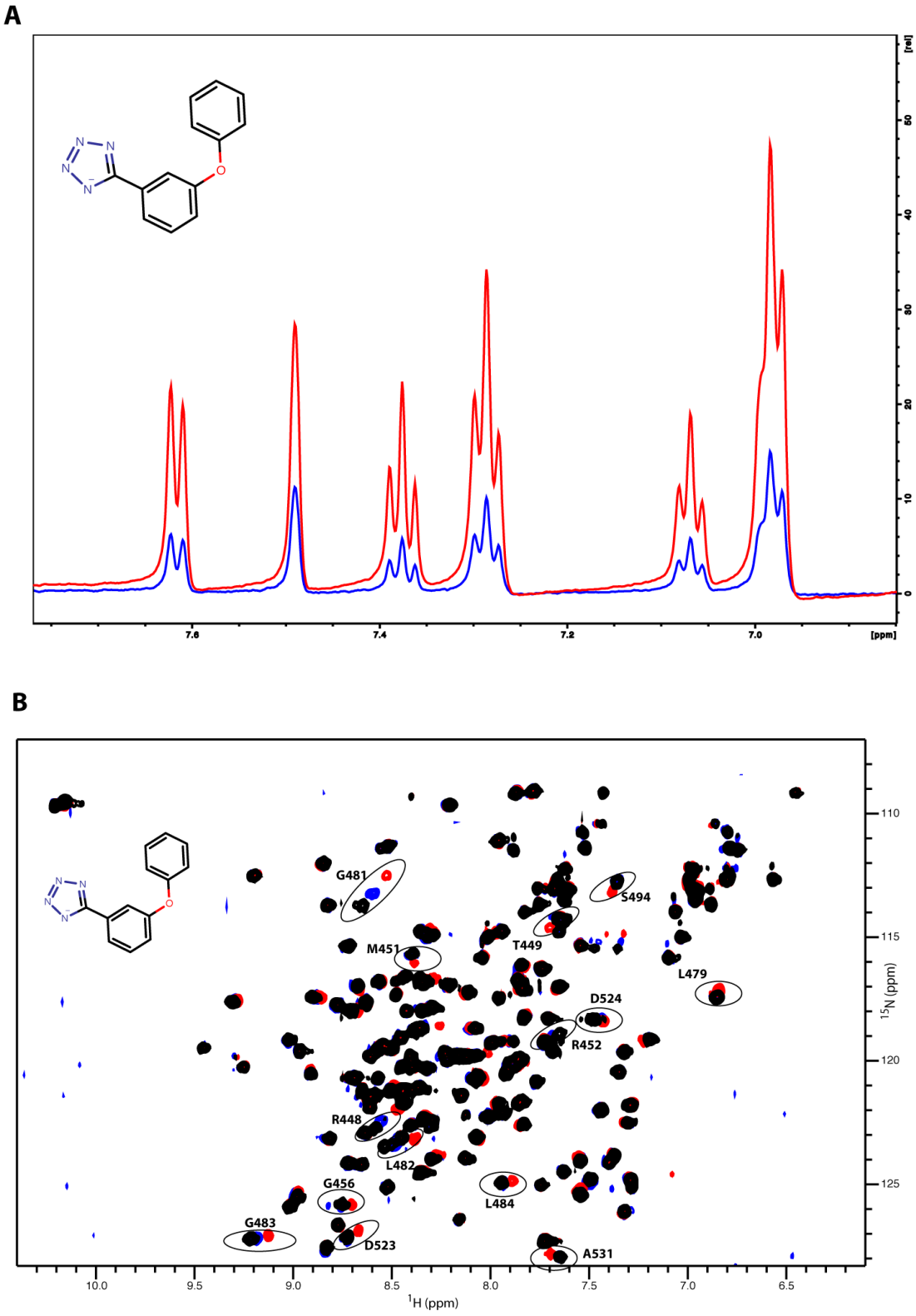


Figure 5.17. (A) STD spectrum of compound CDS001350 using RCD. In red is a 1D ^{1}H reference spectrum, overlayed with a STD spectrum (blue). (B) HSQC spectral overlay of ^{15}N-DnaGC with ZINC72447025 (blue) and CDS001350 (red) at 1 mM respectively. The apo-protein spectrum is show in black.

It was shown that compounds that possess similar binding modes to a protein will alter the same CSPs (Medek *et al.*, 2000). Changes in the position of phenoxy group in

CDS001350, increased the CSP magnitude (shown in the *Appendix C*) and showed a slightly improved the binding affinity of K_D = 1.2 mM as measure by ^{15}N-HSQC titration (*Appendix E*).

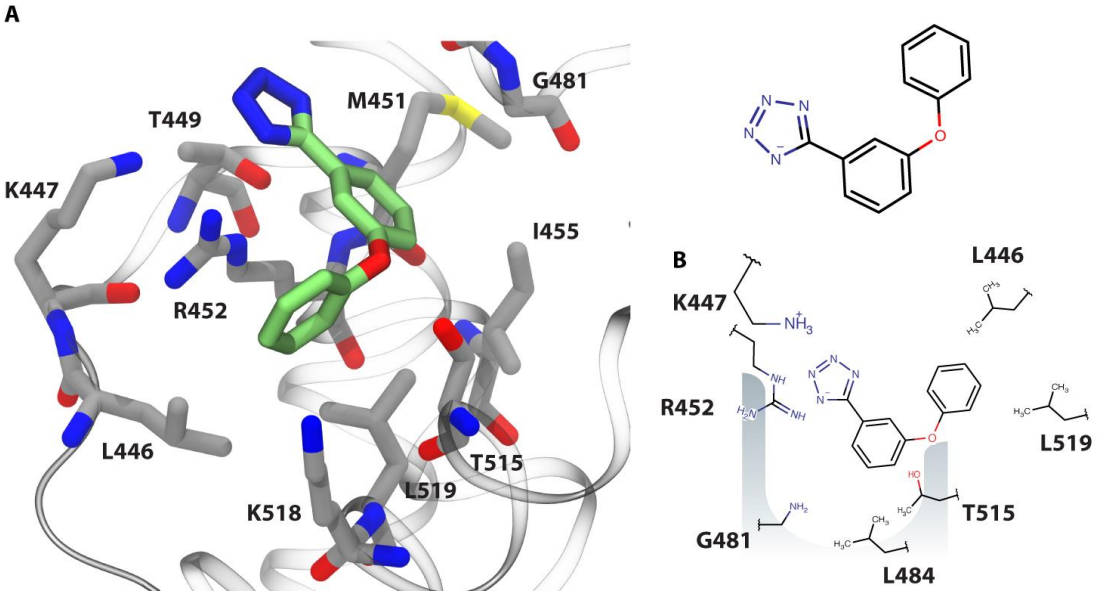


Figure 5.18. Docked binding pose of CDS001350 (A). Schematic representation of interactions (B).

CDS001350 was docked to DnaGC as described previously and is predicted to form electrostatic interactions with the side chains of Lys447, Thr449 and Arg452 (**Figure 5.18**). The central aromatic ring sits in a hydrophobic groove formed by Met451, Leu455 and Leu484, while the oxygen atom from the phenoxy group forms a hydrogen bond with the methyl side chain of Thr515 (**Figure 5.18**). In addition, the aromatic ring of the phenoxy moiety forms a cation-π interaction with the positively charged guanidine group of Arg452, and hydrophobic contact with the Leu446 and Trp522 side chains.

In order to sample the impact of *para*-phenyl position of C4C4 fragment, additional two commercially available compounds were purchased (**Figure 5.19**, compounds CHEMBR7179770 and CHEMBR7693431). Molecular docking suggests that *para*-substitutes dramatically change the orientation of the tetrazole moiety in the pocket. STD, ^{15}N-HSQC experiments confirmed that replacing the fluorine atom with a larger group massively lowers the CSP magnitude. Nonetheless, it was observed even though the *para*-position of phenyl tetrazole is not favoured, addition of chlorine atom slightly increased the CSP magnitude.

In summary the binding of C4C4 tetrazole fragment and related compounds was explored (**Figure 5.19**). *In silico* screening identified ZINC72447025 compound with 3-

fold improved affinity measured by HSQC titration. CDS001350 compound with minimum structural changes showed small improvements in changing the position of phenoxy moiety. The *para*-substituted analogues of C4C4 fragment showed decrease in chemical shifts (and therefore weaker binding), which is most likely caused by reorienting the compound in the binding pocket. Therefore binding affinities were not measured.

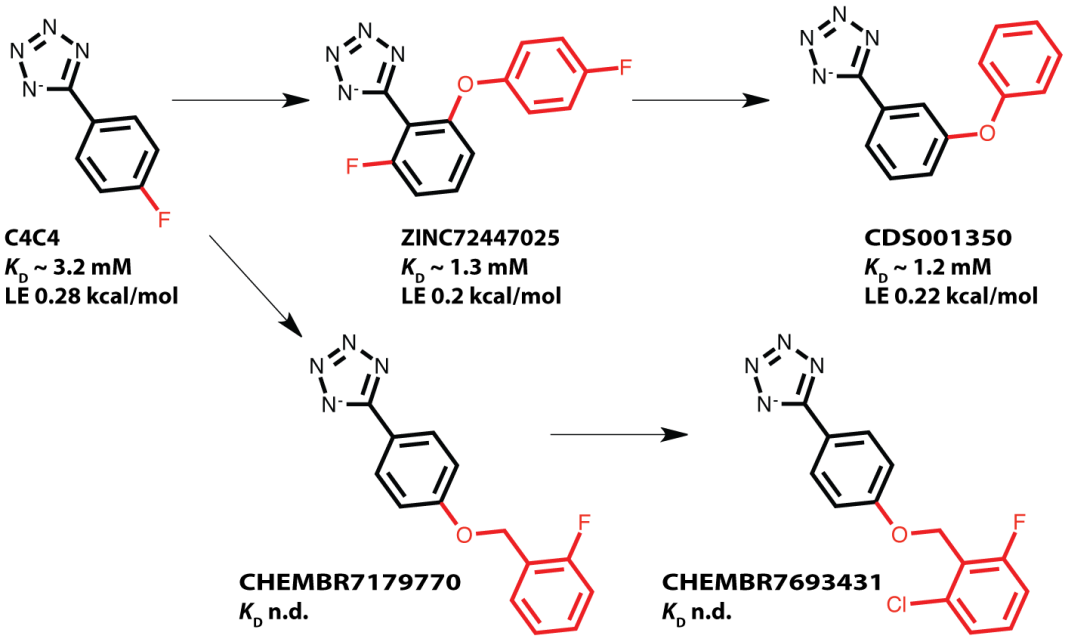
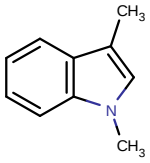
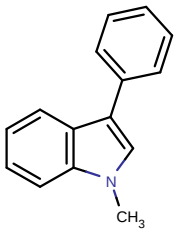
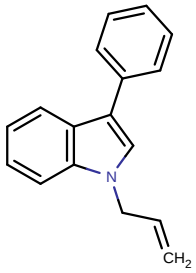
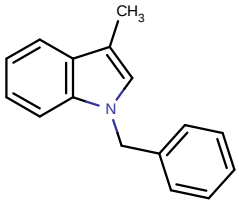
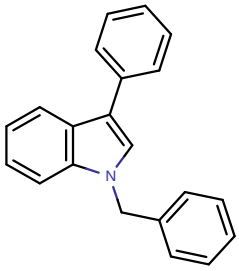
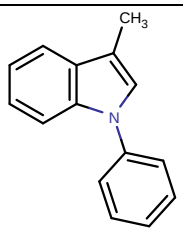


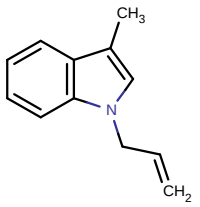
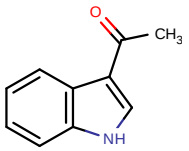
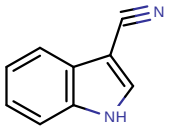
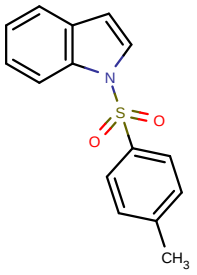
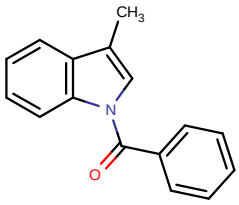
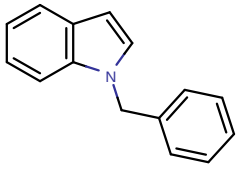
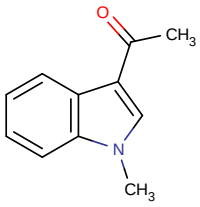
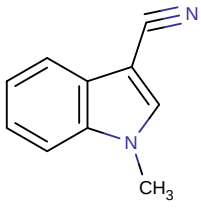
Figure 5.19. Schematic representation of optimisation of C4C4 fragment. The red labelled groups were added along the hit to lead optimisation.

5.2.6 Optimisation of Indole Scaffold

In the first generation hits, the indole scaffold was the most common structural feature. To further explore the indole scaffold, 14 indole analogues with different functional groups were tested (**Table 5.5**) (provided by Dr. Chris Hyland). A similar approach using STD followed by ^{15}N-HSQC experiments confirmed that most of the compounds did not bind, or the CSPs were not significant. This indicates the importance of carboxyl acid groups in the compounds.

Table 5.5. Indole analogues tested for binding to DnaGC by STD and followed up by ^{15}N-HSQC. Y-indicates the binding, N-the absence of binding. Some of the compound did not show any peaks in 1D ^{1}H spectrum as an indicative of low solubility in experimental conditions. Y indicates positive STD while N the absence.

Name	Structure	STD	HSQC-rank
CH1		Y	Rank 1
CH2		-	-
CH3		N	N
CH4		Y	N
CH5		-	-
CH6		Y	N

Name	Structure	STD	HSQC-rank
CH7		N	-
CH8		Y	Rank 1
CH9		Y	Rank 1
CH10		-	-
CH11		-	-
CH12		Y	N
CH13		Y	Rank 1
CH14		Y	Rank 1

5.3 Conclusion

The first generation of hits were investigated by protein-based NMR (^{15}N-HSQC), which also demonstrated that the binding site was identical to the peptide-binding site. The binding affinities were measuring using HSQC titration experiments. For fluorinated compounds such as C4C4, the binding was confirmed by 1D ^{19}F NMR. *In silico* CSP-guided docking approach was utilised to identify likely the binding orientations of hits.

In silico fragment-to-hit approach was utilised for an analogue screen. 7700 commercially available compounds from ZINC database were docked into the SSB-Ct pocket of DnaGC. The compounds with best score were purchased and tested for binding (STD, HSQC). Two tetrazole derivatives with improved affinities were identified. The ZINC72447025 compound was unsuccessfully used to gain the NOE-driven distance restrains.

Chapter 6

Structural Determination of Protein-Ligand Complexes

6.1.1 Introduction

Paramagnetic NMR is an alternative to traditional NOE-based experiments for solving structures (Keizers and Ubbink, 2011; Otting, 2010). Paramagnetic ions are able to weakly align biomolecules along the external magnetic field vector, generating distance restraints (Otting, 2008). By attaching a paramagnetic center to a protein *via* tags makes it possible to apply the method not only to solving structures of protein-protein complexes, but also determining protein structures in complex with ligands.

Paramagnetic entities include organic radicals (e.g. nitroxide radicals) and metal ions (e.g. Mn^{2+}, Gd^{3+}) with unpaired electrons. Paramagnetism is generated as a result of interactions between the unpaired electrons in the 4f orbitals of lanthanides and the nucleus of interest (Su *et al.*, 2006). Organic radicals affect the relaxation properties of proteins (they increase the longitudinal and transverse relaxation rates). The change in relaxation rates between paramagnetic and diamagnetic samples is known as paramagnetic relaxation enhancement (PRE) (**Figure 6.1**). PRE is responsible for line broadening in NMR spectra, and its effect on line width is proportional to the inverse sixth power ($1/r^6$) of distance between paramagnetic centres and nuclei (Otting, 2008). PRE can be used to measure distances of up to 30 Å in proteins and detect conformational changes (Clore *et al.*, 2007).

Paramagnetic lanthanide tags enable the measurement of pseudocontact shifts (PCSs) and residual dipole couplings (RDCs) (Otting, 2008). PCS can be described by the magnetic susceptibility tensor χ defined in terms of three orthogonal axes (χ_x, χ_y, χ_z) (Otting, 2010). If the magnetic moment of the paramagnetic center changes with different orientations of the centre in the magnetic field, the χ tensor is anisotropic. The χ tensor can be decomposed into isotropic and anisotropic components. The anisotropic component ($\Delta\chi$) can be expressed in axial and rhombic components ($\Delta\chi_{ax} = \chi_z - (\chi_x + \chi_y)/2$ and $\Delta\chi_{rh} = \chi_x - \chi_y$). Pseudocontact shifts $\Delta\delta^{PCS}$ arise from those components as follows (**Eq. 6.1**):

$$\Delta\delta^{PCS} = \frac{1}{12\pi r^3}[\Delta\chi_{ax}(3cos^2\theta - 1) + \frac{3}{2}\Delta\chi_{rx}sin^2\theta cos2\varphi]$$

Equation 6.1. $\Delta\chi_{ax}$ and $\Delta_{\chi h}$ are the axial and rhombic component of $\Delta\chi$: r, θ and φ are the polar coordinates of the nuclear spin in respect to the principal axes of $\Delta\chi$ (Bertini *et al.*, 2002).

This relationship has been used to solve solution-state protein structures (Yagi *et al.*, 2013), structures of protein-protein complexes (Hass and Ubbink, 2014) and in drug discovery for solving the structure of small-molecule protein complexes (John *et al.*, 2006; Guan *et al.*, 2013; Pintacuda *et al.*, 2007). In addition to PRE, RDC and PCS, the presence of a paramagnetic centre in a structure results in Curie-dipolar cross-correlated relaxation (CDCCR) effects. While a comprehensive description of RDC and CDCCR are beyond the scope of this thesis, the effects of PRE and PCS shown in **Figure 6.1** can be exploited to characterise structure and function of biomolecules. Paramagnetic NMR was initially used to study proteins with metal-binding sites. Efforts to develop new methods have extended its applicability to include proteins without native metal-binding sites can be studied through introduction (by site-directed mutagenesis) of solvent-exposed cysteine residues and the subsequent covalent attachment of tags containing paramagnetic ions. For readily interpretable PCSs, the paramagnetic tag must be rigid relative to the protein (Shishmarev and Otting, 2013). Flexibility of the tag will reduce the intensity of PCSs and different tag locations may need to be trialled to optimise rigidity.

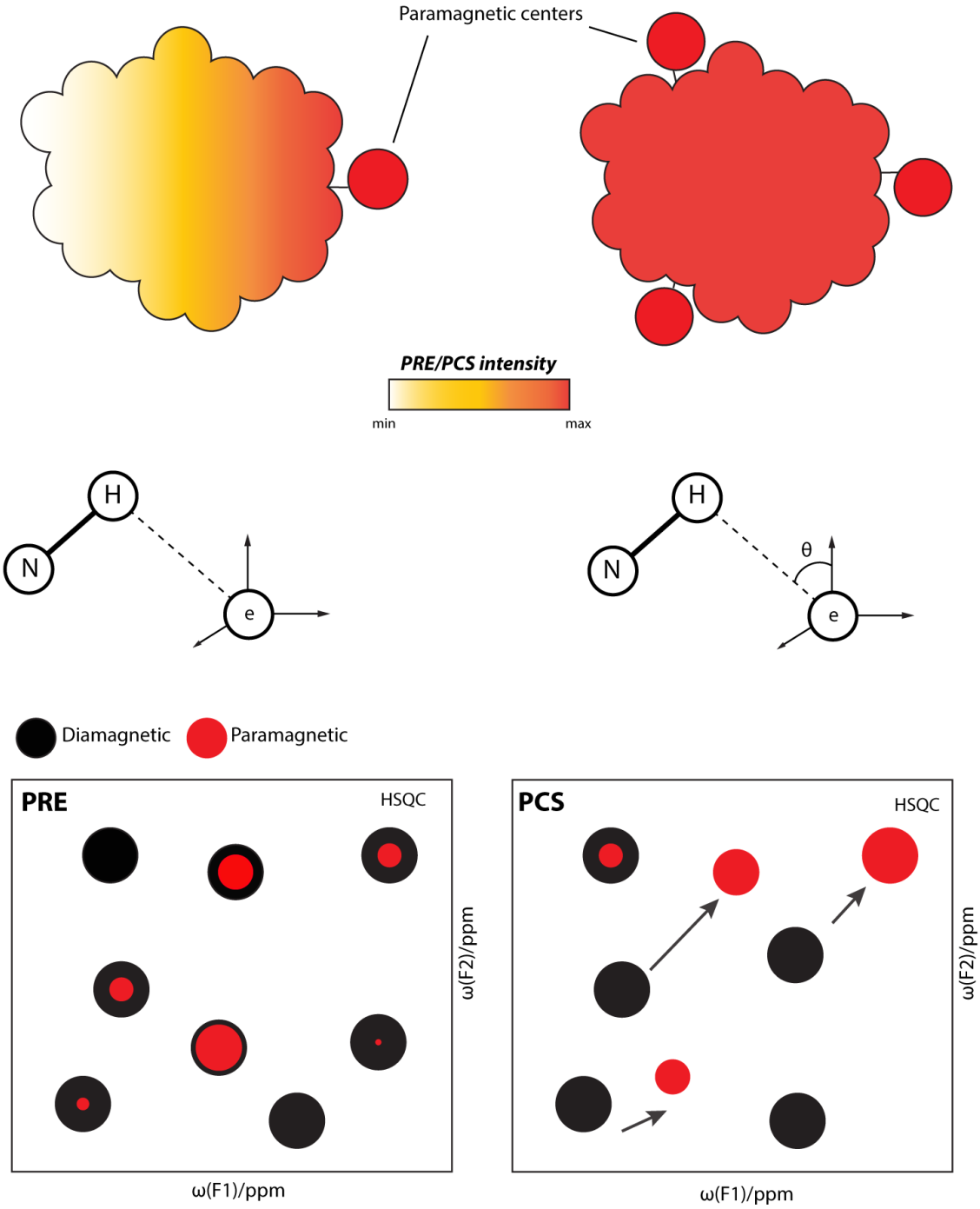


Figure 6.1. Paramagnetic effects caused by presence of tags attached to the protein. Top figure represents the magnitude of measured paramagnetic effect of one and three tags. Bottom figure highlights changes resulted by PRE and PCS in the recorded HSQC spectra. PRE gives peak intensity changes (on the bottom left) whereas the PCS causes chemical shifts and in some cases similar effect to PRE.

The main difference between PRE and PCS is that PRE gives distance restraints from the paramagnetic centre, while PCS possesses additional angular dependence relative to the paramagnetic tag (Koehler and Meiler, 2011). The PCS effect varies by a factor of $1/r^3$ with distance (r) between the paramagnetic centre and the nucleus of interest and

hence can be used to measure long-range distances compared with PRE. It has been shown that, the case of the dysprosium (III) ion, PCSs are capable of measuring distances up to 40 Å from the paramagnetic centre to a nucleus of interest (Allegrozzi *et al.*, 2000; Pintacuda *et al.*, 2007).

6.1.2 Aim

Determination of DnaGC structures in complex with hits using paramagnetic NMR. To test the binding of compounds to other SSB-Ct binding partners (*E. coli* and *A. baumannii* χ subunit of Pol III, PriA and RNAse HI).

6.2 Results and Discussion

6.2.1 Paramagnetic NMR experiments

6.2.1.1 Preparation of lanthanide tagged DnaGC

In order to apply paramagnetic methods, the protein of interest must first have a metal binding site, or alternatively solvent exposed cysteine residues. In order to accurately orient the Δχ tensor, at least three tags are required at different positions relative to targeted binding site (Koehler and Meiler, 2011; Yagi *et al.*, 2013).

Previous studies showed that T499C, E532C and D537C mutants of DnaGC have poor expression levels (Dr. Karin Loscha personal communication), therefore new positions were chosen. The positions of new cysteine residues relative to SSB-Ct binding site is shown in **Figure 6.2**. The position of cysteine mutants were chose to be within 10 to 30 Å distance from SSB-Ct binding pocket of DnaGC. The Q445C, E517C and S545C mutants were cloned (by Dr. Nan Li) and expressed in ^{15}N-rich minimal media as described for ^{15}N-labelled DnaGC in section *2.2.1.7*.

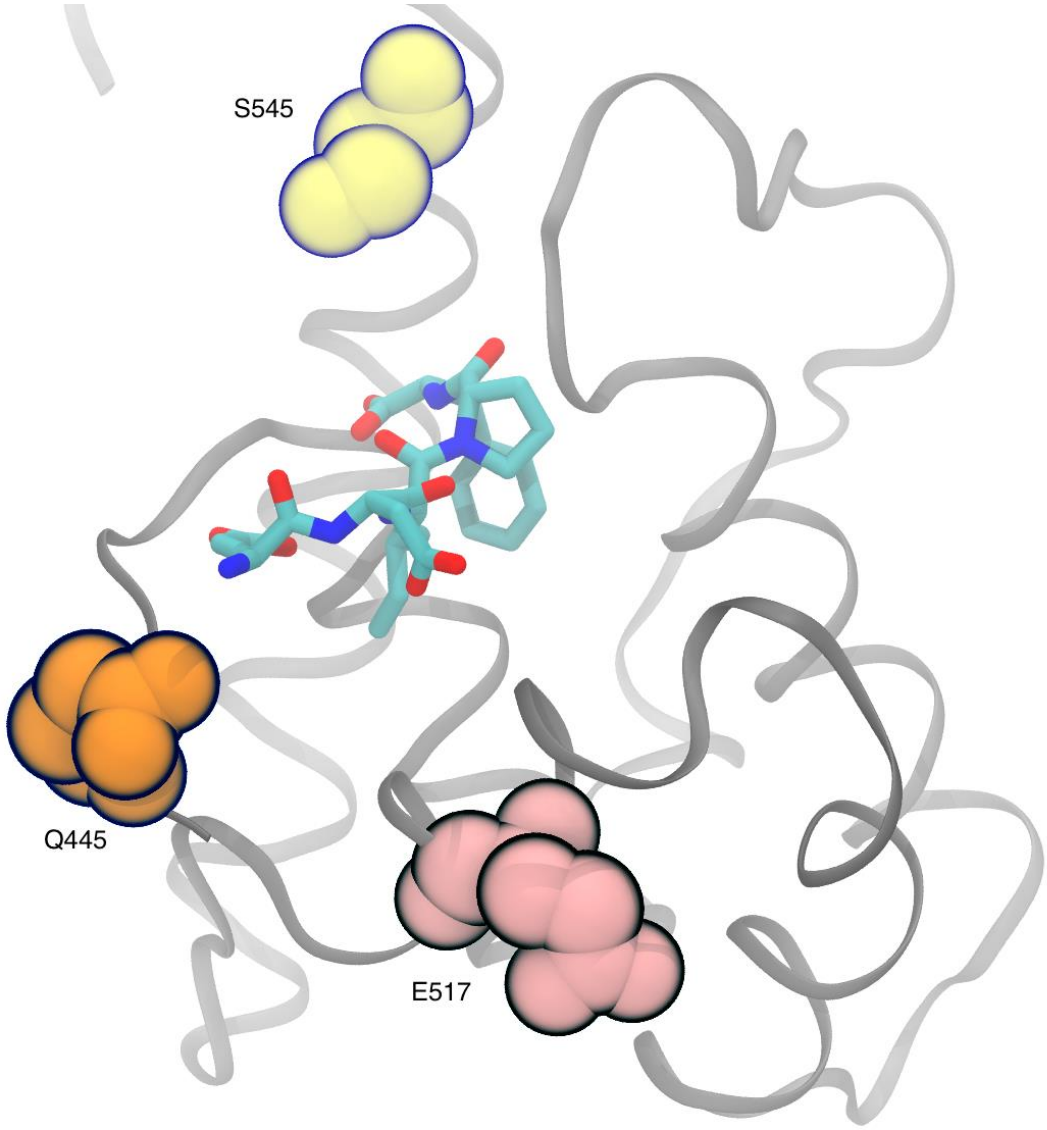


Figure 6.2. Positions of three residues mutated to cysteine relative to SSB-Ct binding pocket.

Unfortunately, purification of the S545C mutant of DnaGC failed, as the protein was inseparable from other *E. coli* proteins. The other two mutants had relatively low expression level. The cysteine mutants were designed such that the introduced cysteine side chains would be solvent exposed and thus would need to be maintained in a reduced state. Therefore, the samples were treated with excess of DTT (5 mM) in order to protect them from oxidisation. Prior to tagging the protein with a paramagnetic complex, the excess DTT was dialyzed out with NMR buffer as described in *section 2.2.4.10*.

The diamagnetic ($C2Y^{3+}$) and paramagnetic tags ($C2Tm^{3+}$, $C2Tb^{3+}$) (**Figure 6.3A**) were applied to two of the DnaGC mutants as described in *section 2.2.4.10*.

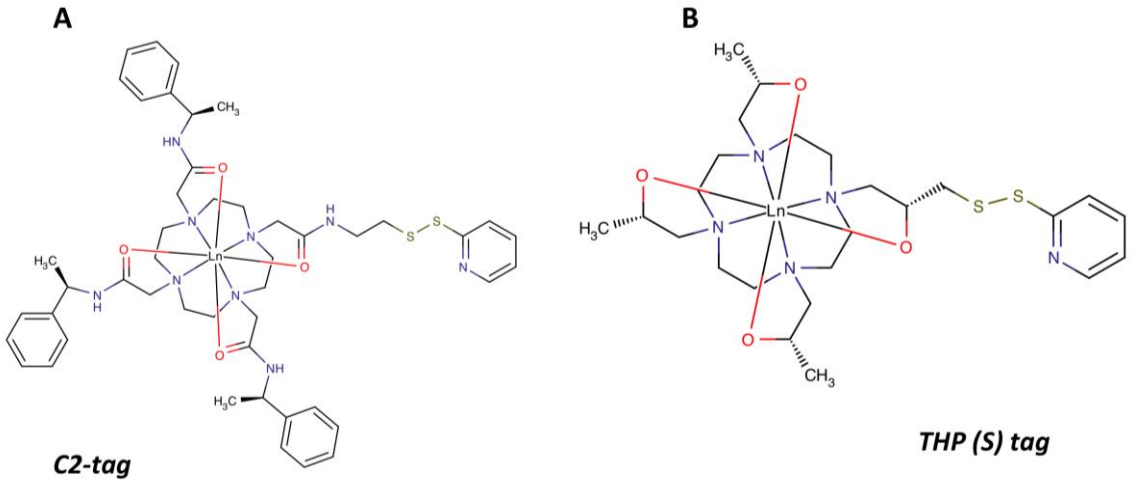


Figure 6.3. Chemical structure of C2 and THP (*S*) tags.

The ligation efficiency of paramagnetic tags is very high, and after overnight incubation, the proteins were expected to be fully ligated.

6.2.1.2 PCS Measurements

Lanthanide induced PCSs are measured by recording ^{15}N-HSQC spectra. ^{15}N-HSQC was recorded for both C2-tagged Q445C and E517C mutants. The Q445C mutant did not produce significant PCSs, possibly as a result of flexibility of the tag, whereas E517C mutant shows distinct PCSs (**Figure 6.4**).

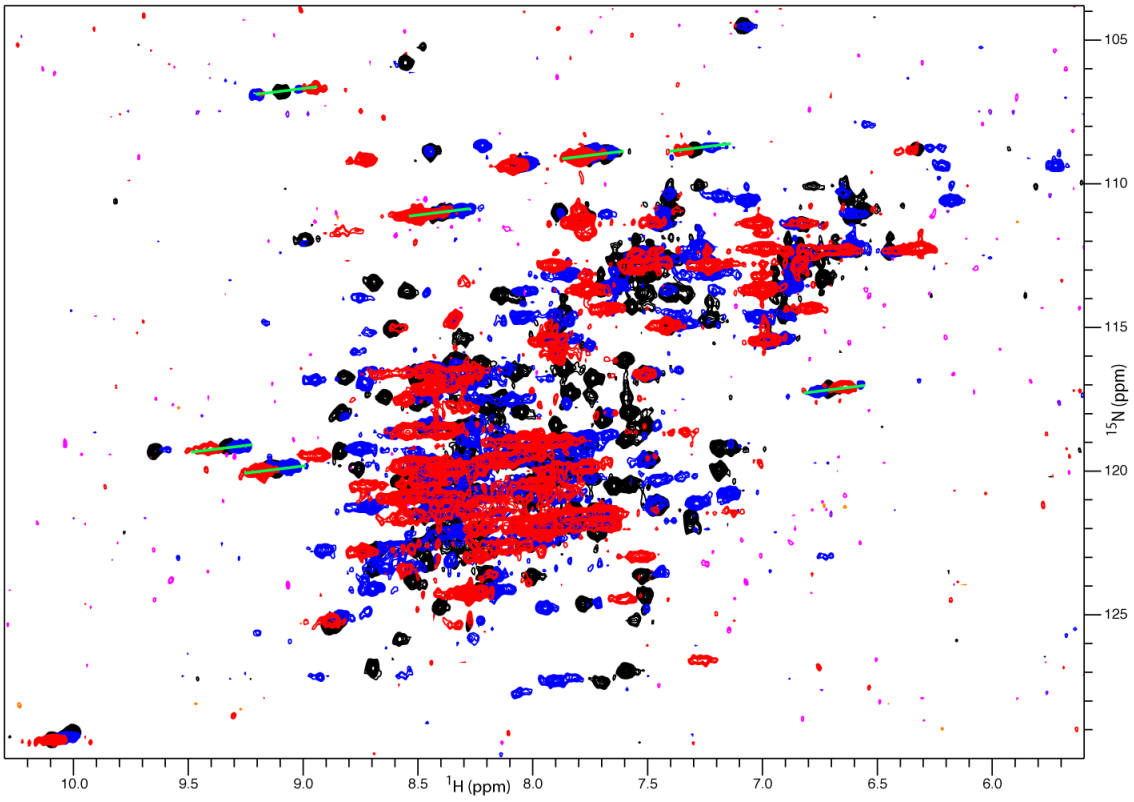


Figure 6.4. ^{15}N-HSQC overlay of C2-tagged E517C mutant containing diamagnetic Y^{3+} (black), Tm^{3+} (blue) and Tb^{3+} (red) respectively.

PCSs were evaluated for only ^{1}H spin to minimise residual anisotropic chemical shifts due to smaller chemical shift anisotropy (John *et al.*, 2005). PCSs were calculated by subtracting the observed diamagnetic chemical shifts from paramagnetic (**Table 6.1**).

Table 6.1. Calculated PCSs of ^{15}N-labelled E517C DnaGC measured with C2 tag. Values are in ppm.

Residue	**Tm^{3+}**	**Tb^{3+}**
Gly438	-0.08	0.09
Arg441	-0.13	0.04
Met451	-0.11	-
Leu457	-0.08	0.1
Leu471	-0.06	0.01
Asn473	0.1	-0.04
Lys478	-0.14	0.01
Leu479	0.05	-0.07
Gly481	0.2	0.2
Leu482	-0.04	0.1

Residue	**Tm^{3+}**	**Tb^{3+}**
Gly483	0.09	-0.15
Leu484	0.17	-
Arg486	0.13	-0.15
Leu504	-0.05	0.06
Gly509	0.11	-0.14
Thr510	-0.02	0.02
Asn512	-0.05	0.06
Thr515	-0.03	-
Asp523	0.24	-0.25
Glu558	-0.06	0.05

Residue	Tm^{3+}	Tb^{3+}
Thr560	-0.03	0.01
Gly562	-0.02	0.02
Ser564	-0.06	0.08

Residue	Tm^{3+}	Tb^{3+}
Asn565	-0.05	0.05
Glu566	-0.08	0.09
Leu574	0.08	-

Numbat software (Schmitz *et al.*, 2008) was used to fit the $\Delta\chi$ tensor to the structure of DnaGC/SSB-Ct (Tak Lo, 2012) using **Eq. 6.1** (Bertini *et al.*, 2002). Substantial spectral overlap in the 7.6-8.6 ppm (^{1}H) region decreased the possibility of unambiguous assignment of the peaks. Nonetheless, about 25 well-resolved peaks were used for calculations.

Subsequently, DQF-COSY, HMBC spectra were acquired and assigned for the CDS001350 compound (*data not shown*). In order to observe PCSs altered by CDS001350 binding to DnaGC, the compound was mixed with C2-tagged E517C mutant in a 4:1 molar ratio and 1D ^{1}H spectra were recorded with diamagnetic and paramagnetic labels (**Figure 6.5**).

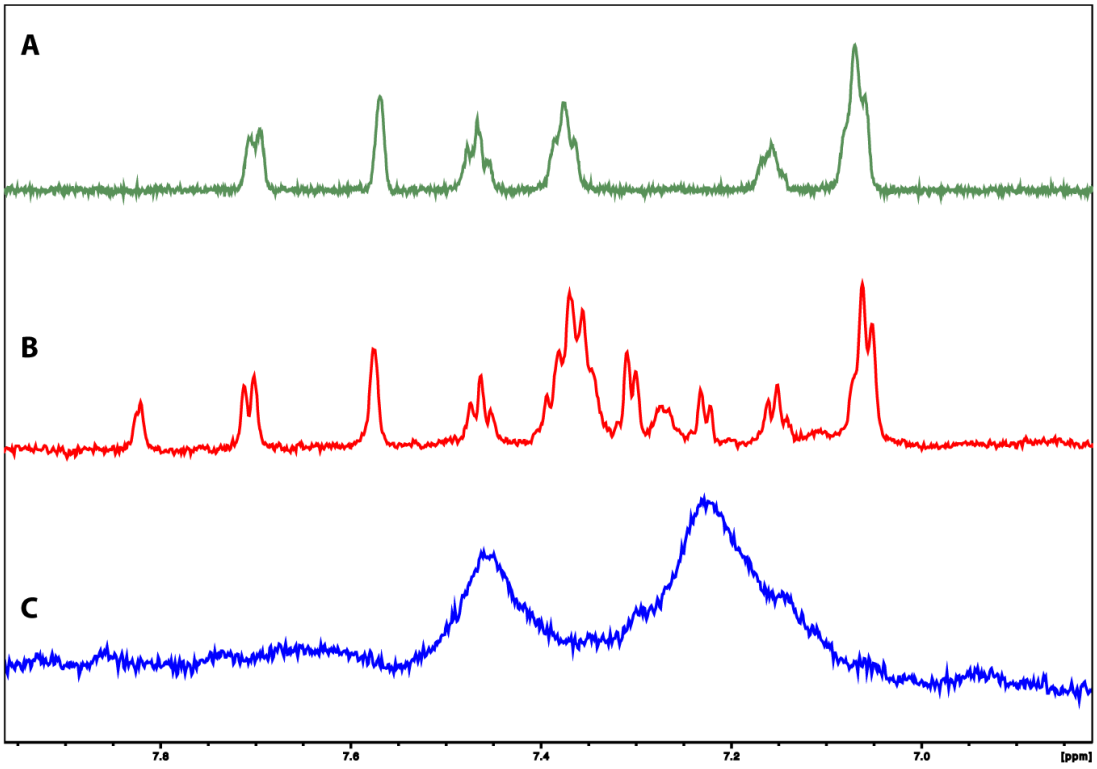


Figure 6.5. 1D ^{1}H spectra of CDS001350 compound at 1 mM (A), with C2Y^{3+}-tagged (B) and C2Tm^{3+}-tagged (C) E517C DnaGC.

1D ^{1}H spectra for the diamagnetic sample gave well-resolved peaks with few additional peaks (**Figure 6.5B**). Extra signals were most likely a result of a residual amount of free tag or other chemical impurities from the tag preparation. 1D ^{1}H spectra were recorded with the same molar ratio of CDS001350 to C2Tm^{3+}-labelled DnaGC. Unfortunately two broad peaks were detected with low spectral resolution (**Figure 6.5C**). Severe line broadening can be explained by a large PRE effect of the C2-tag.

Subsequently, a new tag, THP was tested (**Figure 6.3B**). The (*R*), *(S)*-THP tags were ligated as described for the C2-tag. ^{15}N-HSQC was acquired for (*S*)-THP-tagged E517C mutant of DnaGC. The protein showed PCSs: about 10 peaks were unambiguously assigned (*data not shown*). The same 4:1 ligand-protein ratio with CDS001350 compound was used for measuring PCSs for the compound. To eliminate false positives caused by non-specific tag-inhibitor interactions, a 1D ^{1}H NMR spectra of tag plus CDS001350 at a 1:4 ratio was recorded (**Figure 6.6**).

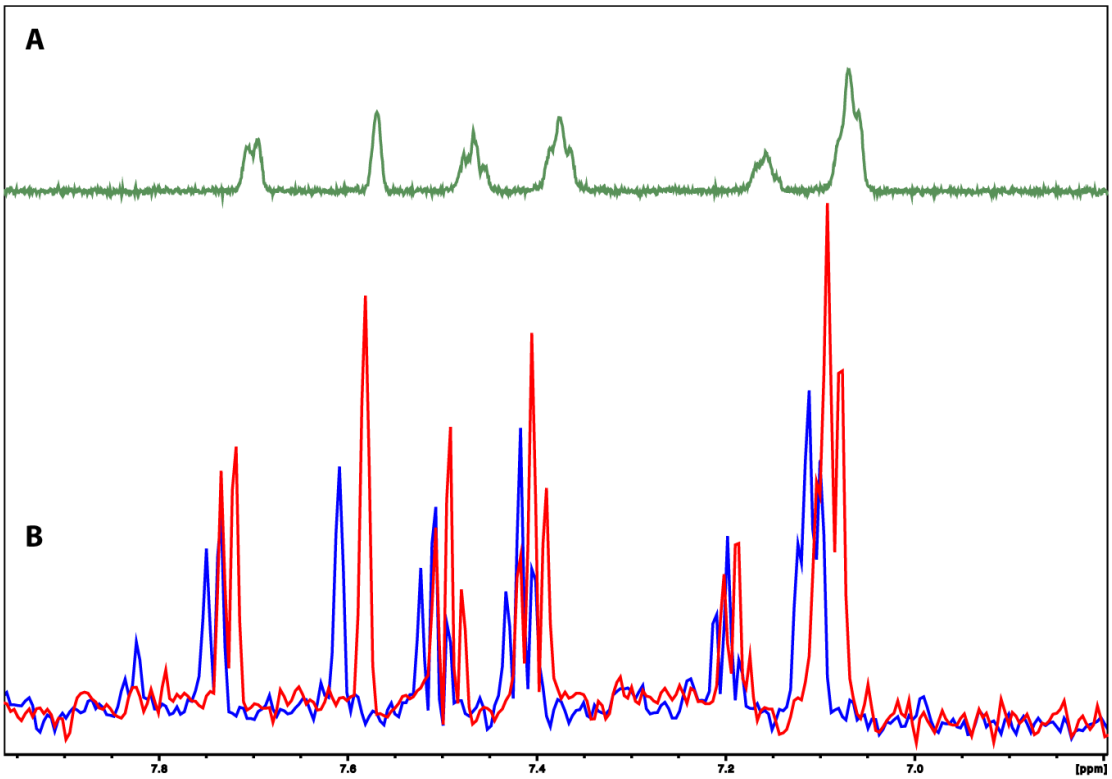


Figure 6.6. 1D ^{1}H spectra showing CDS001350/(*S*)-THP tag interaction. (A) 1D ^{1}H spectrum (green) of CDS001350 compound at 1 mM, (B) 1:4 ratio mixture of CDS00135/(*S*)-THP. Blue spectrum represents paramagnetic (Tm^{3+}), red is diamagnetic (Y^{3+}) respectively.

Surprisingly, CDS001350 in the presence of (*S*)-THP-tag showed undesirable non-specific interaction, demonstrated by comparison of chemical shifts in 1D ^{1}H NMR spectra (**Figure 6.6**). Due to the interactions of tag with inhibitor, it was not possible to

use PCS for structural calculations. The most likely explanation for this interaction is the electrostatic attraction between the positively charged tag and the negatively charged tetrazole ring. The lack of negatively charged tags is a current limitation of the method.

6.2.1.3 Paramagnetic Relaxation Enhancement

The Q445C mutation is located relatively close to the SSB-Ct binding pocket, enabling the possibility of PRE-NMR experiments. The residues close to the tag, including the ligand, will experience a strong paramagnetic effect even if the life time of the state is short (Volkov and Worrall, 2006). The most common method for utilizing the PRE effect is through the introduction of a nitroxide spin label. MTSL is the most popular such label, as it has high selectivity for cysteine thiols.

The Q445C mutant was labelled with MTSL (described in *2.2.4.10*) and 1D ^{1}H spectra were recorded in presence of CDS001350 compound at 1:4 and 1:6 molar ratio with MTSL (paramagnetic) and reduced MTS (diamagnetic) samples (**Figure 6.7**).

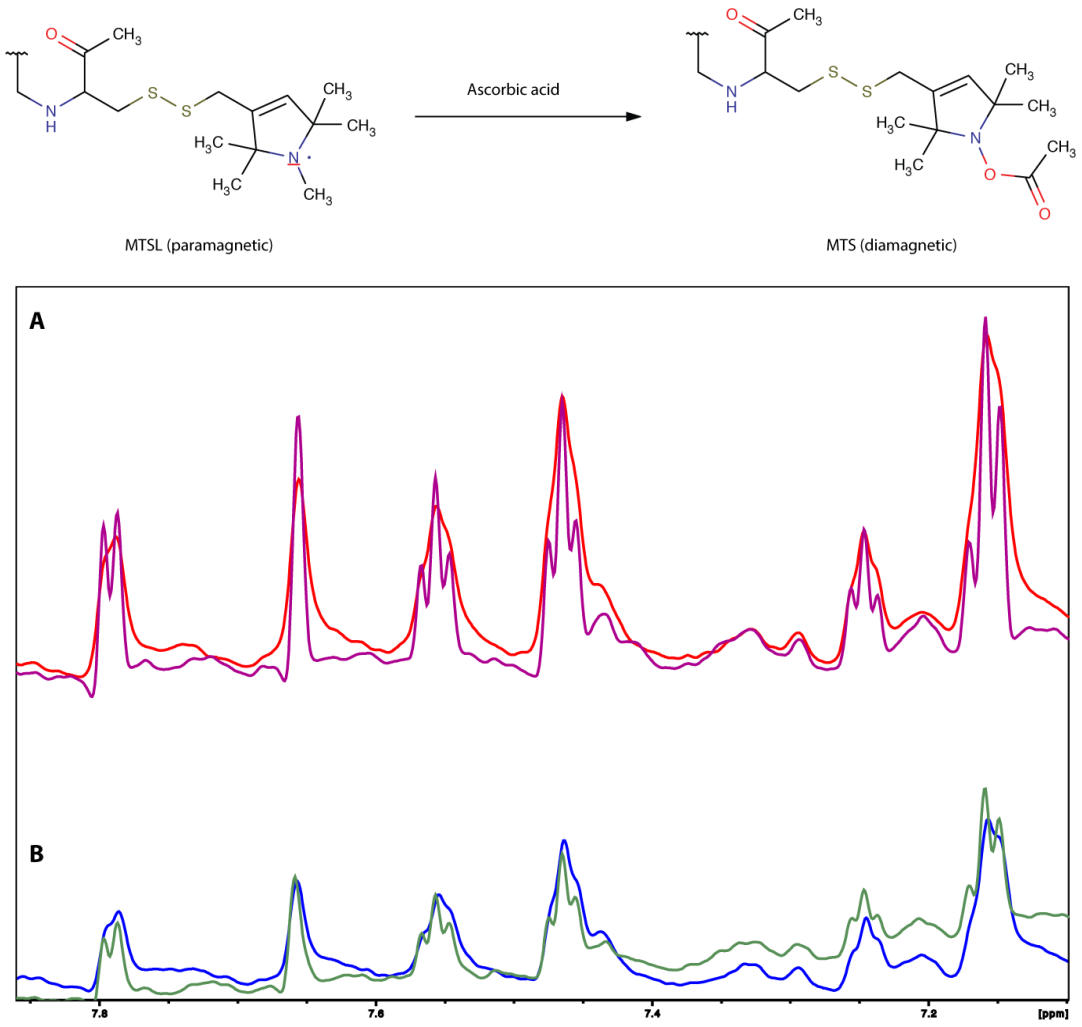


Figure 6.7. At top is shown the reduction reaction of MTSL to MTS. Below are shown 1D ^{1}H PRE measurements with MTSL labelled Q445C DnaGC mutant with CDS001350 compound at 1:6 (A) and 1:4 (B) molar ratio. Red and blue spectra indicate paramagnetic, green and magenta diamagnetic samples respectively.

Two different concentrations of MTSL/MTS labelled samples were used in an attempt to exclude random intermolecular interactions by showing a consistent PRE effect. However, in both cases, line broadening to the same extent was observed (**Figure 6.7**). The PRE effect was evenly distributed for all CDS001350 signals, thus makes PRE did not yield useful information. The limited precision of PRE can be explained by the flexibility of the MTSL group and the small molecular size of the inhibitor compared to the protein.

6.2.2 Phenylalanine moiety as a starting point

As described earlier, SSB is a protein interaction hub known to bind more than 14 different proteins involved in various stages of replication, repair (Shereda *et al.*, 2008). One such protein is ExoI, 3' to 5' exonuclease. Recently the crystal structures of compounds disrupting the ExoI/SSB-Ct interaction were identified using a HTS approach: (Z)-2-(5-(3-bromobezylidene)-4-oxo-2-thioxothiazolidin-3-yl)-3-phenylpropanic acid (BOTP), (E)-2-(5-(2-methyl-3-phenylallyl)-4-oxo-2-thioxothiazolidin-3-yl)-2-phenylacetic acid (MPTA), 2-((2-chloro-5-(trifluromethyl)phenyl)amino)-5-methoxybenzoic acid (CFAM) and 5-(tert-butyl)-2-(6-chlorobenzo[d]thiazole-2-yl)-2,4-dihydro-3*H*-pyrazol-3-one (BCBP) (**Figure 6.8**) (Lu *et al.*, 2010).

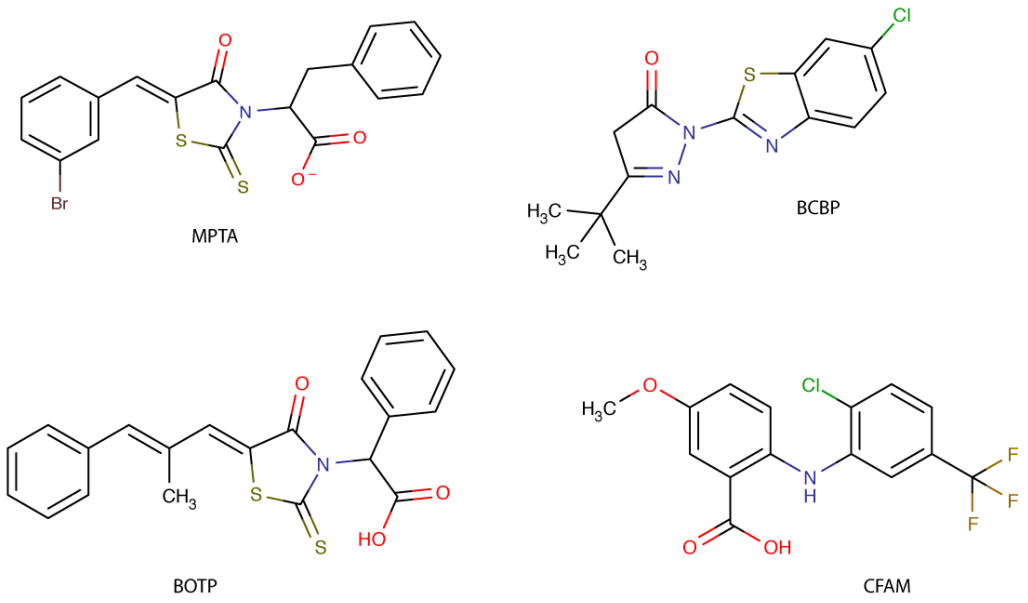


Figure 6.8. Chemical structures of inhibitors of interactions with the C-terminus of SSB.

Furthermore, the inhibitors were tested with other SSB-Ct binding partners (RecQ, PriA) and binding was also observed (Lu *et al.*, 2010). FP competition assays demonstrate that these compounds bind competitively to DnaGC in competition with SSB-Ct. The binding of these compounds was determined using fluorescent polarisation (FP) competition assay (Luzuriaga, 2011). The crystal structures of ExoI with bound inhibitors revealed that part of the compounds mimic the Phe177 residue of SSB-Ct (Lu

et al., 2010). In order to study the interaction with DnaGC, 2D ^{15}N-HSQC NMR experiment was utilised (**Figure 6.9**).

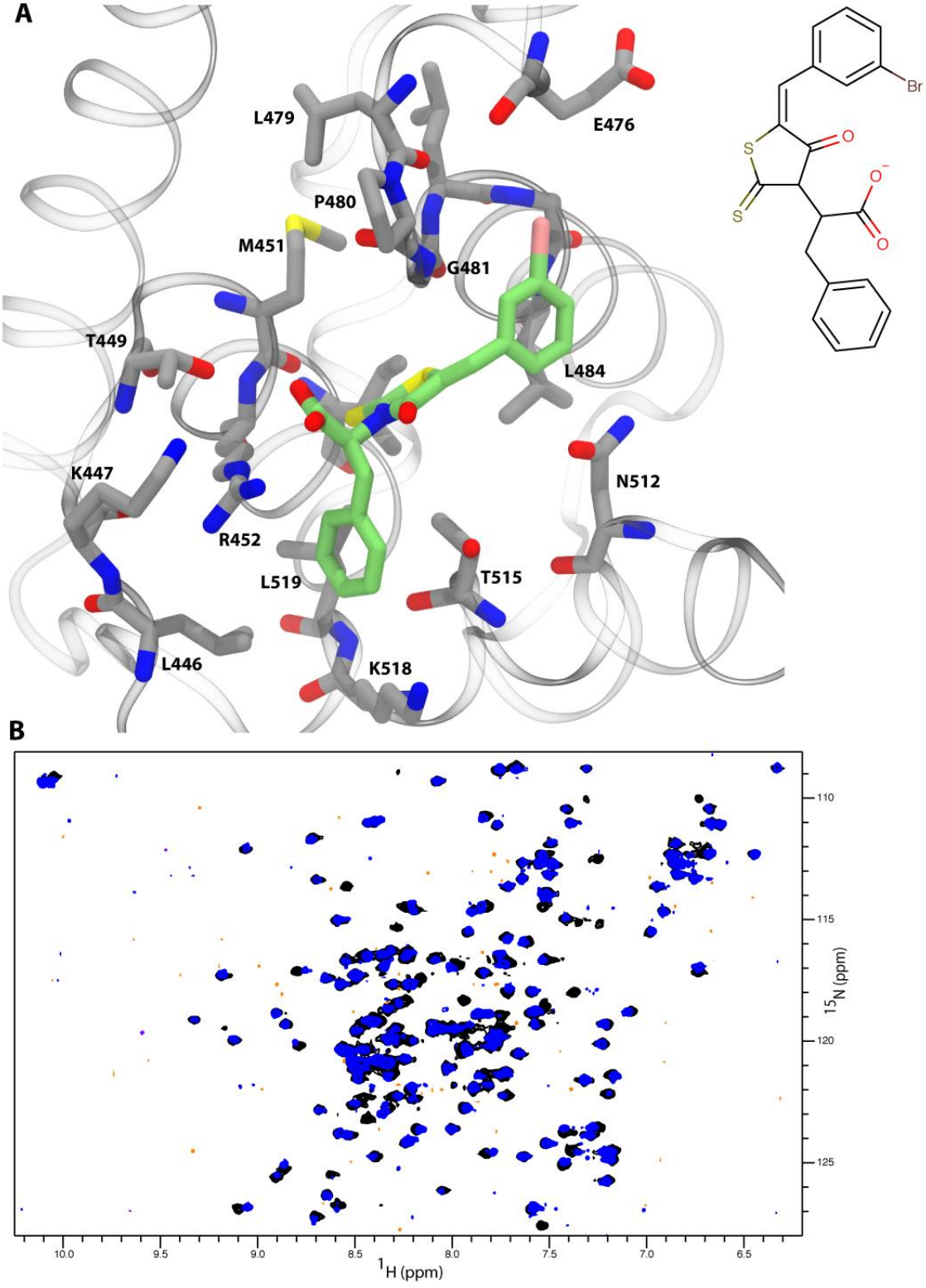


Figure 6.9. BOTP docked into DnaGC. (A) Lowest energy binding pose of BOTP in complex with DnaGC, generated by docking. (B) The spectral overlay of apo-protein (black) and in the presence of 1mM BOTP (blue).

Using the method described in *Chapter 5*, BOTP (1 mM) was added to ^{15}N-labelled DnaGC and 2D ^{15}N-HSQC spectrum was recorded. Addition of the compound result in loss of peaks, demonstrating that the interaction was in intermediate exchange (Williamson, 2013).

In all currently available crystal structures of SSB-Ct with binding partners, and in MD simulations described in *Chapter 3*, the side chain Phe177 forms hydrophobic interactions with the various binding pockets while the carboxylic acid motif forms strong electrostatic/hydrogen bond interaction with a basic residue. The possibility of using phenylalanine as a starting point for fragment-to-hit optimisation was therefore considered. Based on experimental and *in silico* data, N-Acetyl-L-phenylalanine was tested using STD and by a subsequent ^{15}N-HSQC experiment. Both experiments confirmed binding, although CSP magnitudes compared to other fragment was significantly weaker, inferring that binding affinity is weaker. ^{15}N-HSQC titration experiments were conducted. The measure binding affinity was about 2-fold weaker compare to the fragment hits (K_D ~ 6 mM). Similar to the method described in *Chapter 5*, the analogue screen performed by molecular docking using ZINC Database identified N-Phenylacetyl-L-phenylalanine, which was tested (**Figure 6.10**).

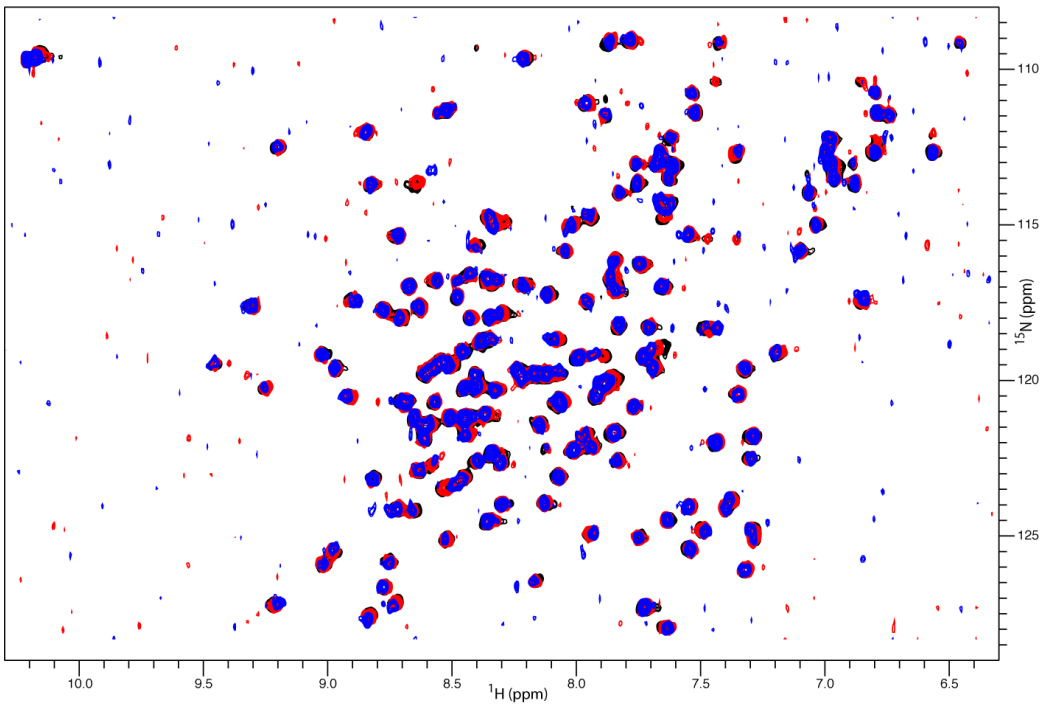


Figure 6.10. Spectral overlay of DnaGC with N-A-L-Phe and N-PheA-L-Phe. 2D spectra for the apo-protein (black), protein plus N-A-L-Phe (red) and protein plus N-APhe-L-Phe (blue) are shown.

The ^{15}N-HSQC spectra for N-Phenylacetyl-L-Phenylalanine showed greater CSP-magnitudes compare with N-Acetyl-L-phenylalanine, although no dramatic changes were detected.

Observations made here complemented with available structural data indicating that phenylalanine derivatives could be used to develop inhibitors of SSB-Ct interactions.

6.2.3. Binding of DnaGC Inhibitors to Other SSB-Ct Binding Partners.

In order to explore the possibility that hits against DnaGC could bind to other SSB-Ct-binding partners, fragment hits were tested for binding to SSB-Ct binding partners from *E. coli* and *A. baumannii*. These were *E. coli* Pol III χ, PriA, RNAse HI and *A. baumannii* Pol III χ (*E. coli* χ subunit of Pol III, PriA, RNAse HI were generously provided by Dr. S. Jergic and Dr. N. Li). Binding was assayed using STD-NMR (**Figure 6.11**).

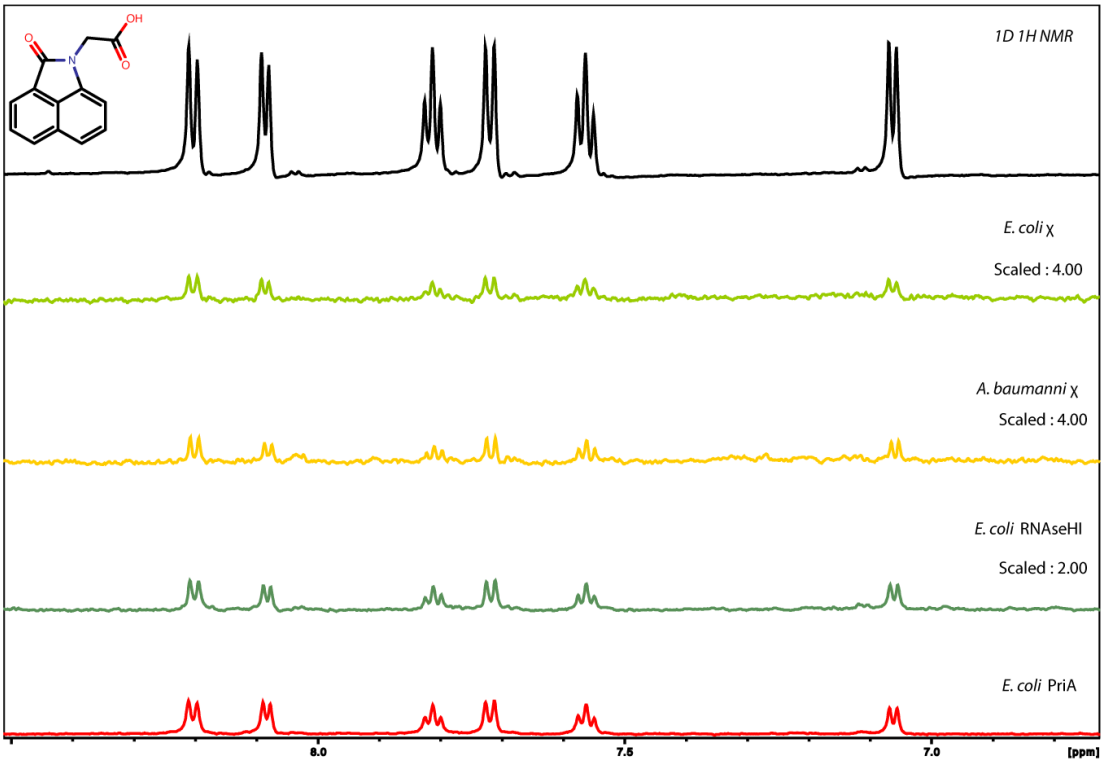


Figure 6.11. STD-NMR of L1G8 fragment with other SSB-Ct binding partners represented in different colours: black-1D reference spectrum, green-*E. coli* Pol III χ, yellow-*A. baumanni* Pol III χ, dark green-RNAse HI and red-PriA respectively.

Remarkably, STD experiments showed binding of four fragments (L1G8, L1H10, C4C4 and L1A11) to the four SSB-Ct binding partners considered here. Additionally, the first generation of tetrazoles were tested and binding was confirmed (*Appendix F*). The best

fragments hits were docked into *E. coli* Pol III χ using the method described in *2.2.4.8.* The resulting binding free energies are shown in **Table 6.2**. The predicted binding orientations of C4C4 and ZINC72447025 to *E. coli* Pol III χ are shown in **Figure 6.12**.

Table 6.2. The binding free energy of docked poses of first generation of fragments generated by Autodock Vina for *E. coli* χ.

	Pose 1	**Pose 2**	**Pose 3**	**Pose 4**	**Pose 5**	**Pose 6**	**Pose 7**	**Pose 8**	**Pose 9**	**Pose 10**
C4C4	-5.9	-5.9	-5.8	-5.5	-5.5	-5.5	-5.5	-5.4	-5.4	-5.4
L1G8	-6.4	-6.4	-6.2	-6.1	-6.1	-6.1	-6.0	-6.0	-5.9	-5.8
L1H10	-5.7	-5.7	-5.6	-5.5	-5.5	-5.5	-5.4	-5.3	-5.3	-5.3
L1A11	-5.2	-5.	-5.2	-5.2	-5.1	-5.1	-5.1	-5.1	-5.1	-5.1
ZINC72447025	-6.8	-6.8	-6.6	-6.4	-6.4	-6.4	-6.4	-6.3	-6.3	-6.2
CDS001350	-6.7	-6.7	-6.7	-6.7	-6.7	-6.6	-6.6	-6.8	-6.5	-6.5

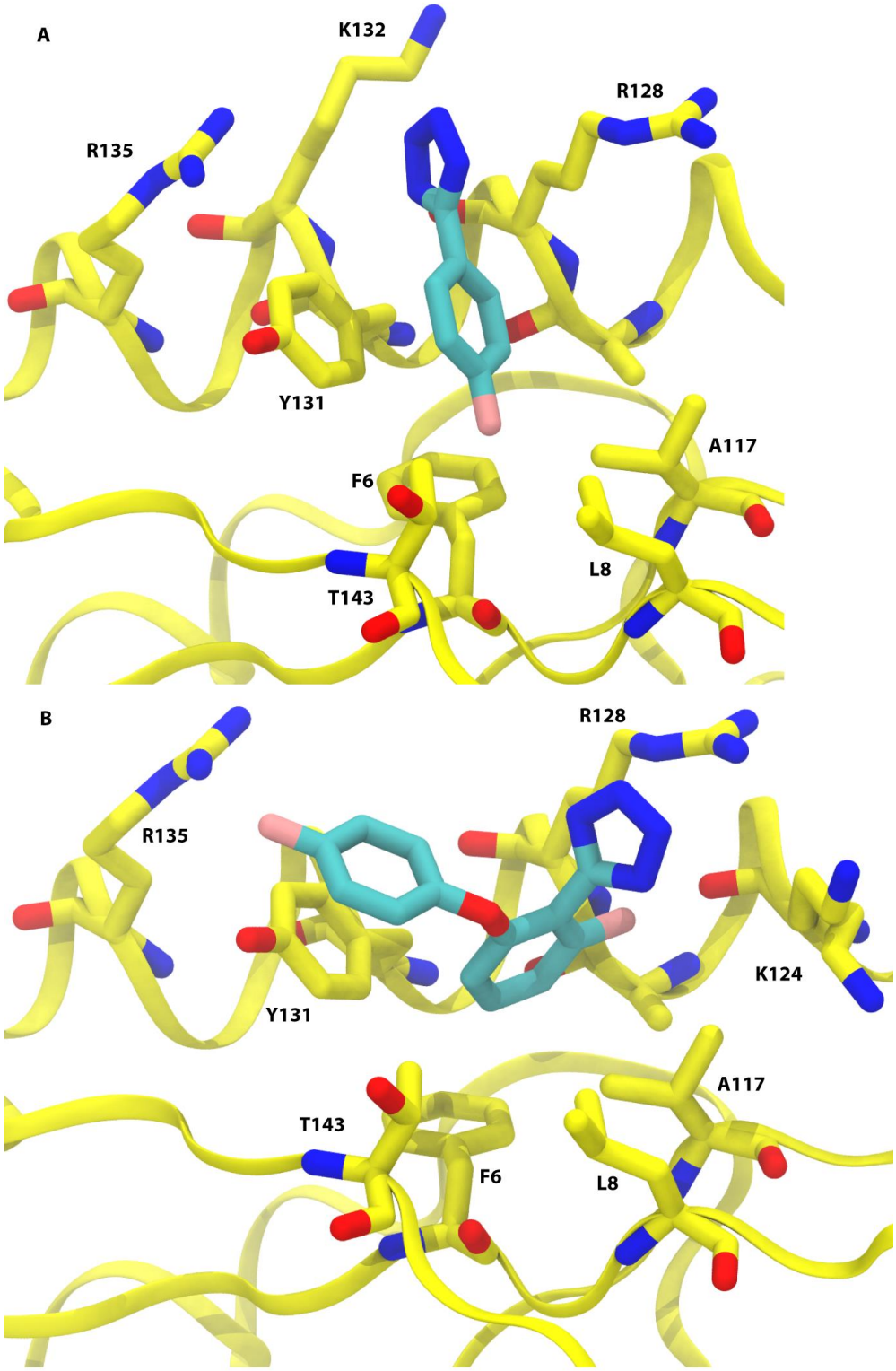


Figure 6.12. Lowest energy orientation of C4C4 (A) and ZINC72447025 (B) compounds with *E. coli* χ. Protein is shown in ribbon coloured in yellow. Residues within 5 Å away from ligand are represented in stick.

For the docked *E. coli* χ/C4C4 complex, three basic residues (Arg128, Lys132 and Arg135) are predicted to interact with the negatively charged tetrazole group, while the phenyl group forms a π-π stacking interaction with Tyr131. The fluorine atom of C4C4 interacts with the side chains of Leu8, Ala117 and Thr143 (**Figure 6.12A**). In case of ZINC72447025, the position of the tetrazole group is rotated about 45° in the model (**Figure 6.12B**). Small rearrangements in the binding pose bring the tetrazole group closer to Arg128 and Lys124. The fluorine atom in the *ortho*-position makes contact with backbone carbonyl group of Lys124. The phenyl ring forms a σ-π interaction with the methyl groups of Leu8 and Ala117. The phenyl ring of the extended phenoxy moiety forms a π-π stacking interaction with Tyr131, positioning the fluorine in close proximity to the hydroxyl group of Tyr131 and side chain of Arg135. The docked models suggest that the fluorine atoms in both C4C4 and ZINC72447025 compounds are involved in various interactions. 1D ^{19}F NMR would agree with the hypothesis based on model. 1D ^{19}F spectra recorded in the presence of a 20 molar excess of ligand were acquired with *E. coli* χ. In 1D ^{19}F spectra of C4C4 fluorine peak broadened and shifted slightly upfield (**Figure 6.13A**).

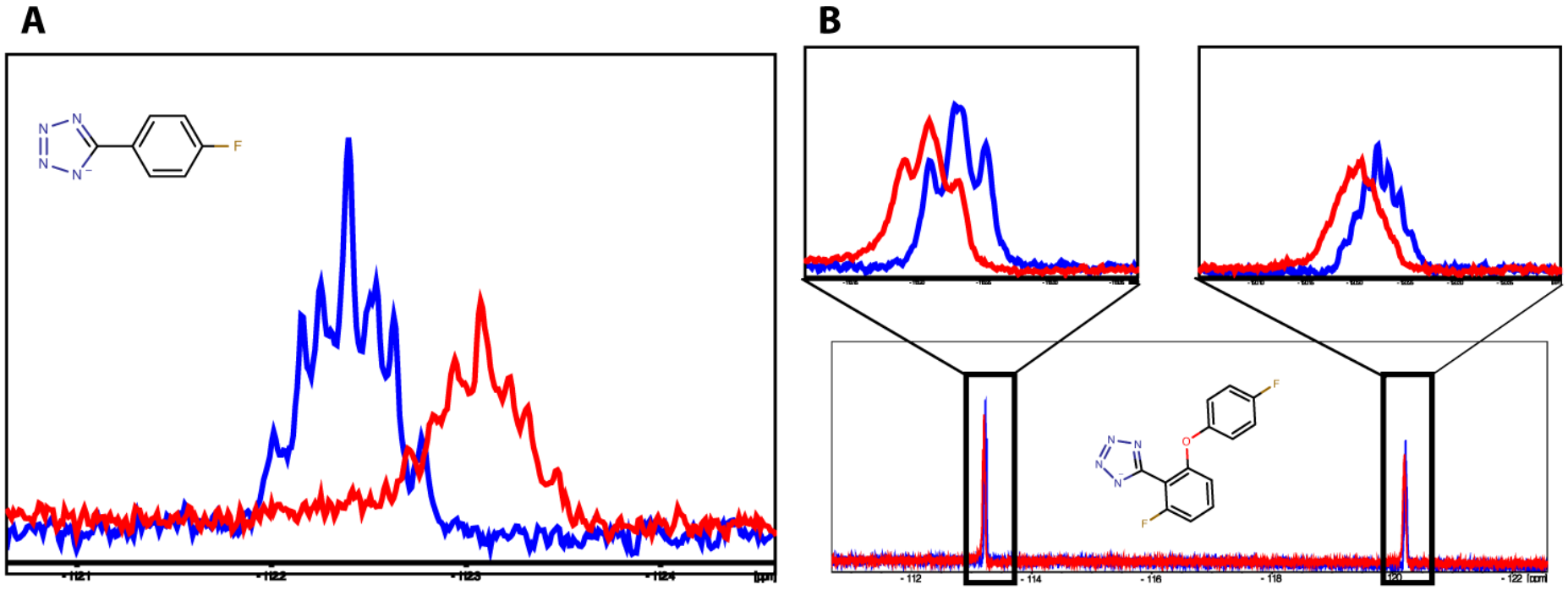


Figure 6.13. ^{19}F-NMR of C4C4 (A) and ZINC72447025 (B) compound in isolation (blue) and with *E. coli* χ in presence of protein at 1:20 protein-ligand ratio (red).

1D ^{19}F experiments were conducted under the same experimental conditions for ZINC72447025/*E. coli* χ and showed similar downfield peak shifts and intensity changes as for DnaGC. Experimentally, these results support the predicted binding models.

A comparison of binding poses of all three tetrazole derivatives with *E. coli* χ suggests that the position of the tetrazole moiety may be different in ZINC72447025, moving away from Lys132 and Arg135 towards Lys124, with Arg128 maintaining its interactions. With respect to compound CDS001350, changing the position of the phenoxy-group is predicted to affect the orientation of the compound in the pocket, with the tetrazole returning to a position similar to that observed for C4C4 (**Figure 6.14**).

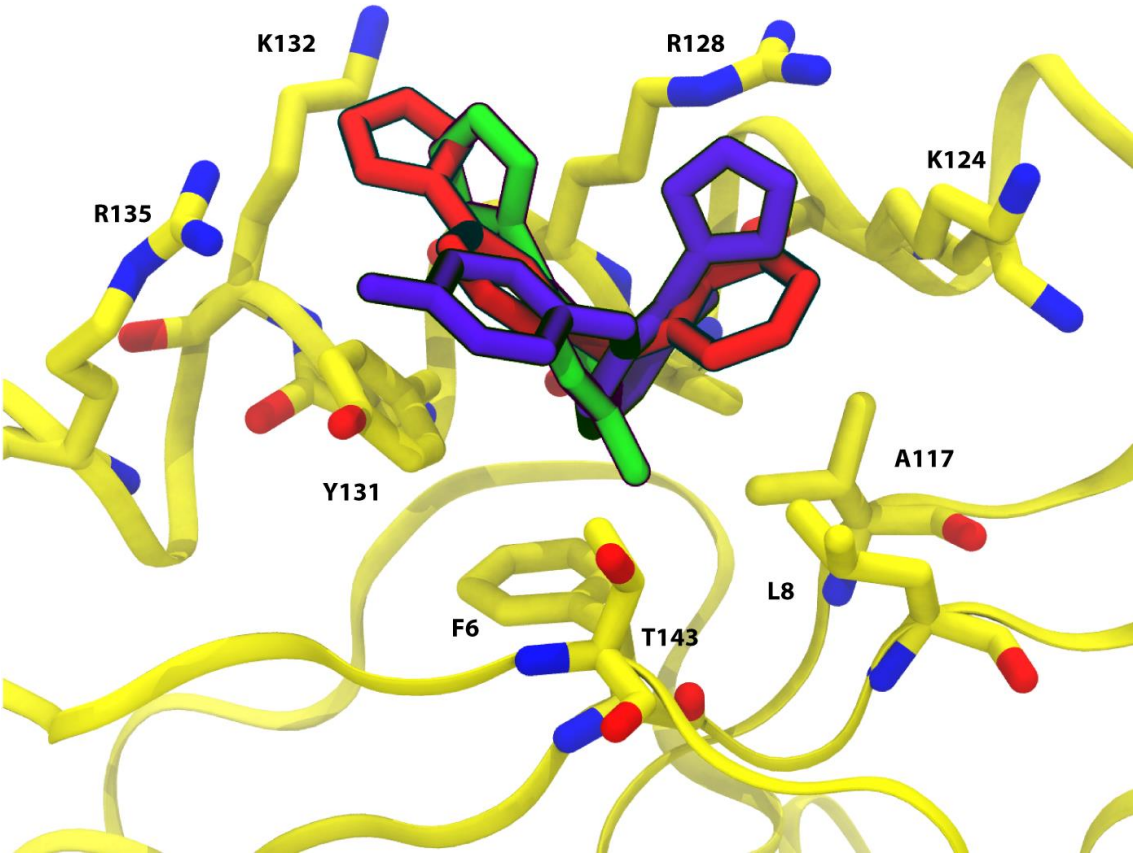


Figure 6.14. Comparison of modelled orientations of C4C4 fragment and two best tetrazole analogues in the SSB-Ct binding pocket of *E. coli* χ. Protein shown in yellow ribbon, C4C4 in green, ZINC72447025 is in violet and CDS001350 fragment is shown in red.

6.3 Conclusion

Paramagnetic NMR was attempted as an alternative method for obtaining distance restraints to characterise for protein/ligand interactions. A PCS-based approach was

taken. Unfortunately non-specific compound-tag interactions were observed. PRE NMR did show changes in relaxation of the fragment, however, these were evenly distributed across the fragment signals, and thus did not yield useful spatial data. Intriguingly, compounds did show binding to other SSB-Ct binding partners: *E. coli* Pol III χ, PriA, RNAse HI and *A. baumannii* Pol III χ.

Chapter 7

Conclusion and Future Directions

7.1 Conclusion

The identification of new classes of broad-spectrum antibiotics acting *via* new targets is of great importance. The rise of so-called "superbugs", i.e. antibiotic resistant infection-causing bacteria-poses a potentially catastrophic threat to public health. The rise of multi- and pan-resistance coincides with an "innovation gap": only two new classes of broad-spectrum antibiotics have been approved in the last 40 years: the oxazolidinones and the lipopeptide antibiotic daptomycin. Almost all classes of modern antibiotics are derived from chemical scaffolds based mostly on natural products discovered over 40 years (Cole, 2014).

The DNA replication machinery is a rich source of targets for antibiotic discovery (Sanyal and Doig, 2012; Robinson *et al.*, 2012). SSB is known to interact with more than 14 binding partners through its C-terminus (Shereda *et al.*, 2008; Tak Lo, 2012; Ryzhikov *et al.*, 2011; Lu and Keck, 2008; Marceau *et al.*, 2011). Simultaneously disrupting interactions of the SSB C-terminus with multiple binding partners has the potential to neutralise bacterial pathogens. The human SSB C-terminus is distinctive and therefore inhibitors of bacterial SSB-Ct interactions are unlikely to react with human SSB-Ct-binding partners (Robinson *et al.*, 2010).

The SSB-Ct can be described as a "short linear motif" or SLiM. SLiMs are short, conserved motifs involved in many vital processes in the cell (e.g. cell proliferation, signalling). The structures of SLiMs are generally not well characterised mainly due to their disordered nature. SLiMs are often in transient contact with their binding partners. Upon binding, some of them become structured. SLiMs have been found to play an essential role in pathogens (e.g. virus host cell interactions, they tend to mimic host cell SLiMs) (Kadaveru *et al.*, 2008; Davey *et al.*, 2011; Yang, 2012). Advances have been made targeting SLiM-mediated interactions in recent years (Wells and McClendon, 2007). Some of these compounds are in clinical trials. For example, Cilengitide (a cyclized Arg-Gly-Glu-containing pentapeptide) targets a SLiM binding site in the treatment of gliomas. Cilengitide selectively blocks binding of integrins αvβ3 and αvβ5 to Arg-Gly-Asp motifs in the extracellular matrix. Inhibition prevents signalling and causing the disassembly of the cytoskeleton in glioma cell (Scaringi *et al.*, 2012). A second example is Nutlin-3, an inhibitor of the MDM2/p53 interaction (Secchiero *et al.*, 2011). MDM2 binds a short hydrophobic motif in the N-terminal domain of p53, and

nutlin-3 blocks the SLiM binding site on MDM2, leading to p53 stabilization and activation of cell cycle arrest and apoptosis in cancer cells.

Currently there are four crystal structures of SSB-Ct (or segments of it) in complex with its binding partners: DnaGC (Tak Lo, 2012), RecO (Ryzhikov *et al.*, 2011), ExoI (Lu and Keck, 2008) and the Pol III χ/ψ subunits (Marceau *et al.*, 2011). In all structures N-terminii of the SSB-Ct peptide has high B-factor, and was not detected in electron density. Molecular dynamic simulations of four SSB-Ct binding partners DnaGC, RecO, Exo I and Pol III χ/ψ subunit in complex with SSB-Ct were computed. The simulations shed light on similar features of SSB-binding pockets and in the dynamic behaviour of bound SSB-Ct. Critical features that could be exploited in drug design were identified. It was observed in all systems that the three hydrophobic residues (the IPF sub-motif) at the C-terminus were relatively rigid, while the three contiguous aspartate residues (the DDD sub-motif) were flexible. Positively charged residues were identified on the surface of binding partners that engaged in electrostatic interactions with the aspartate residues of SSB-Ct. All crystallographically observed interactions were maintained in the MD simulations. In all complexes considered, a salt bridge occurs between the SSB-C terminal carboxylic acid and an arginine side chain. The impact of these positively charged resides were in agreement with mutagenesis data. The local dynamics of SSB-Ct was assessed by residue-based RMSD and order parameter calculations. The SSB-Ct peptides showed fluctuations that increased with distance from the C-terminus. These were in agreement with weighted atomic density and order parameter calculations, indicating that in all cases, the three contiguous aspartate residues do not bind in a specific pocket and are disordered. The IPF sub-motif is more stable, rigidly bound and engages in specific interactions. Clustering analysis indicates that there are highly populated clusters in the DnaGC and RecO simulations, and low-population clusters in ExoI and χ/ψ simulations. *In silico* mutagenesis of Ile175Leu and three aspartate residues to asparagine (Asp173Asn, Asp174Asn and Asp174Asn) in SSB-Ct have shown that the loss of charged residues in the DDD sub-motif destabilizes SSB-Ct-binding. These along with existing mutagenesis studies emphasise the importance acidic residues in the SSB-Ct peptide.

One particular SSB-Ct binding partner, DnaG primase was subjected to fragment screens to find fragments that could act as starting points for antibiotic development. Initially, attempts were made to use X-ray crystallography for fragment screening. However, good quality crystals of neither DnaGC (nor RCD) were obtained. NMR and

SPR-based methods were therefore utilised. Fragment-based screening against DnaG primase targeting the DnaGC/SSB-Ct interaction was successful. Two fragment libraries were screened against DnaG primase using various biophysical techniques (summarised in **Figure 7.1**).

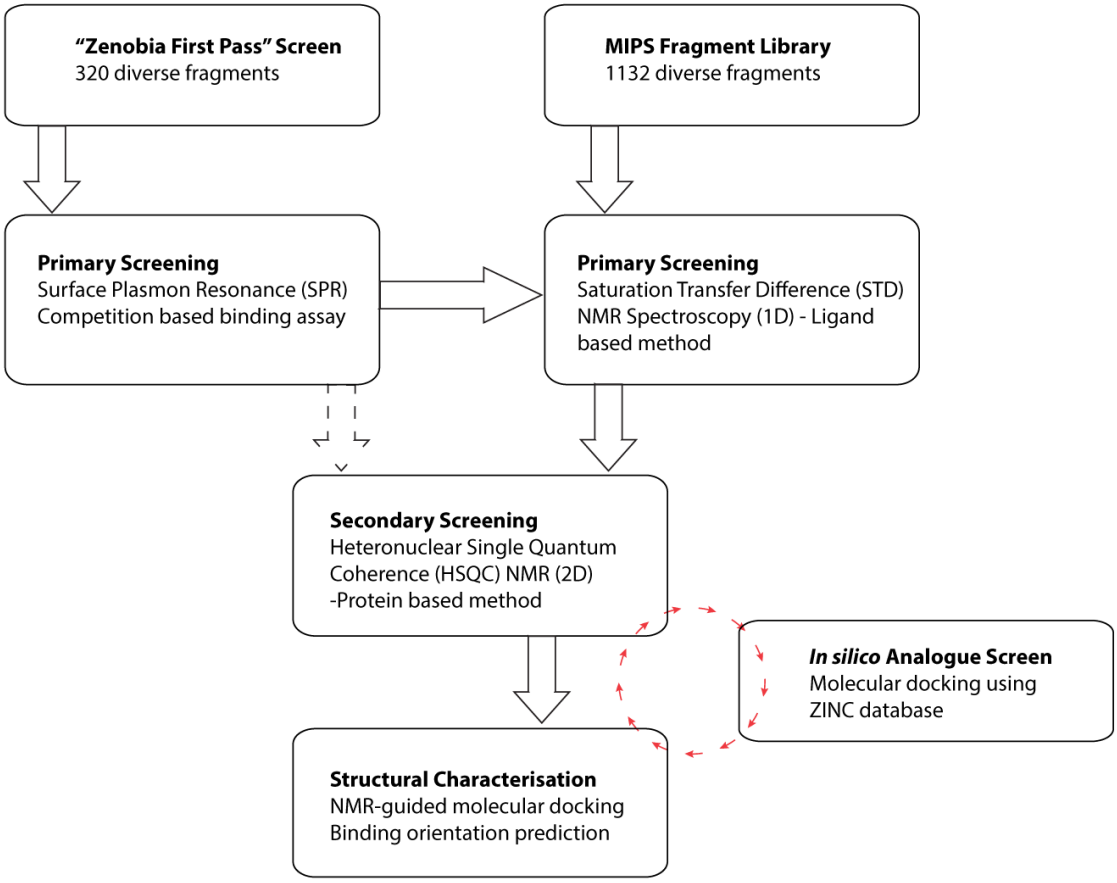


Figure 7.1. Summary of screening workflow. Two fragment libraries were used for primary screening. Identified hits were confirmed by ^{15}N-HSQC.

The Zenobia Therapeutics "first pass screen" was screened using a SPR competition assay. Fragments competed with immobilized biotinylated SSB-Ct peptide on the SPR chip surface for the peptide-binding pocket on DnaG primase. Six compounds (A9, B10, C4, D6, D8 and C1) were identified that competed for occupying the SSB-Ct binding site on DnaGC. SPR hits were validated by STD-NMR. A second fragment library, the Monash Institute of Pharmaceutical Science fragment library (1132 fragments), was screened against RCD using STD-NMR. The STD signal intensity was used to rank the compounds. STD-NMR gave 17% hit rate (200 fragments). The high hit rate can in part be explained by non-specific binding to the protein surface, compound aggregation in the cocktail mixtures and the presence of two domains in the

RCD construct provides a greater surface area for fragments to bind to. STD hits were ranked according to their signal intensity.

Fragments known to be non-specific binders ("frequent hits") were excluded from the candidate compound list. 2D ^{1}H-^{15}N HSQC was used to filter the remaining rank 3 and 2 hits for those binding to DnaGC. Some optimisation was carried out to improve the HSQC experimental conditions, modifying the buffer that was used in all measurements. CSP magnitude was used to rank the compounds. Based on STD signal intensities and CSP magnitude changes in the HSQC experiments, hits from the MIPS library were more potent compare to SPR hits. The carboxylic acid group was common functional group in identified fragments. It was shown for some of the compounds (e.g. L1C6) and indole scaffold that the presence of a carboxylic acid moiety in the compounds is necessary.

The binding orientations of hits were generated by experiment-driven (CSP-guided) molecular docking using Autodock Vina 1.1.2. Thus predicted binding poses were in agreement with experimentally observed chemical shift perturbations. For fluorine-containing ligands, 1D ^{19}F-NMR was employed. Peak shift and intensity changes were observed upon addition of DnaGC to the ligand, providing further evidence in support of the predicted binding poses. The predicted binding site of the fragments is in occupying the IPF-binding pocket of DnaGC.

Five fragments (C4C4, L1G8, L1H10, L1A11 and L-phenylalanine) were selected for fragment-to-hit optimisation, all having millimolar binding affinity to DnaGC. In all five fragments there is a carboxylic acid motif (except C4C4, which contains a tetrazole group) and hydrophobic aromatic ring. Binding affinities were estimated using ^{15}N-HSQC titration. High throughput *in silico* docking was used to screen the ZINC library for commercially available analogues of the fragment hits. Overall 7700 compounds were docked into the DnaGC pocket. 30 analogues of first generation of hits were purchased for testing the binding to DnaGC. A series of phenoxy tetrazole derivatives were identified with 2-fold improvement in binding affinities: ZINC72447025 (K_D 1.3 mM) and CDS001350 (K_D 1.2 mM). Calculated binding orientations suggest that Arg452 is crucial to the binding of the tetrazole moiety *via* a salt bridge interaction. 3D ^{15}N NOESY-HSQC was carried out on one such compound, ZINC72447025 with the aim of generating models of the bound compound guided by NOE-based distance restraints. Due to the high micromolar affinity compound, no protein to ligand NOEs were observed.

Paramagnetic NMR methods were then explored as an alternative approach to obtain distance restrains. Three DnaGC cysteine mutants (Q445C, E517C and S545C) were designed and cloned. Two (Q445C, E517C) were successfully expressed and purified. Two different experiments were carried out monitoring PCSs and PRE changes. For the PCS measurements, two lanthanide tags (C2 and (*S*)-THP) were tested. The C2 tag caused extensive line broadening therefore (*S*)-THP was used. PCSs for ^{15}N-DnaGC and CDS001350 compound were observed, however, non-specific tag-compound interactions contributed to overall PCSs, rendering the data unusable. Non-specific tag-compound interactions may be attributed to electrostatic attraction between the positive charges on the tag and negatively charged tetrazole moiety in the compound.

Paramagnetic relaxation enhancement experiments were attempted as a last resort. MTSL-paramagnetic tag was attached to Q445C mutant of ^{15}N-DnaGC for PRE experiments. Two concentrations of MTSL-labelled DnaGC with CDS001350 compound were used to measure the intensity changes for paramagnetic (MTSL) and diamagnetic (MTS) samples, however, the PRE effects on the compound signals were evenly distributed. This is likely due to high flexibility in the MTSL tag.

Five of the first generation hits, along with two optimised tetrazole derivatives were tested for binding to four other SSB-Ct binding partners (*E. coli* χ, PriA, RNAse HI and *A. baumannii* χ) using ligand-based NMR experiments. All tested compounds have shown binding to similar extent to DnaGC.

7.2 Future Directions

The results presented here provide insights into SSB-Ct interactions and SSB-Ct binding sites as drug targets. Based on the similar patterns behaviour of SSB-Ct in MD simulations presented in Chapter 3, the design of multi-target inhibitors of SSB-Ct binding may be possible. HTS approaches have identified inhibitors of the ExoI/SSB-Ct interaction (Lu *et al.*, 2009). Four compounds were identified (BOTP, MPTA, BCBP and CFAM) that have a phenylalanine-like motif in their structure and bind to RecO, PriA. However further experiments conducted on the effects of these compounds in *Bacillus subtilis* species was due to off-target activity, affecting protein synthesis (Marceau *et al.*, 2013).

MD simulations carried out in this work indicated that SSB-Ct binding partners have similar structural features. Therefore, compounds identified by Lu and co-workers could be co-crystallised with other SSB-Ct binding partners to elucidate the three-dimensional

structure of the complexes, that could help in structure-guided optimisation of compounds for multi-target binding.

A major drawback in this project was the lack of robust crystals for determination of three-dimensional structures of DnaGC/ligand complexes. Future work could explore the crystallisation of homologues of DnaGC and/or RCD from different species other than *E. coli*. Recently, the crystal structure of the DnaG C-terminal domain from *Vibrio cholera* was deposited in the PDB (PDB ID 4IM9; Abdul Rehman, Tarique and Gourinath). This protein shows 42% sequence identity with respect to *E. coli* DnaGC, the SSB-Ct binding site is conserved, and the SSB-Ct binding site is not blocked by crystal contacts. It should be noted that the C-terminal 6 residues of *V. cholera* SSB are identical to those in *E. coli*, i.e. DDDIPF. Once SSB-binding to *V. cholera* DnaGC is confirmed, this crystal form could be useful for determining structures of complexes with the compounds reported here. Also, new cysteine mutants could be produced for further paramagnetic NMR experiments. New paramagnetic tags (e.g. CLaNP-5) (Keizers *et al.*, 2008) are becoming available that have been proven to be more reliable compare to single-armed tags.

The compounds described in this thesis that bind to DnaGC represent a starting point for the development of tighter-binding ligands. This may be achieved by optimisation of the tetrazole scaffold. This could involve extending the best compound (CDS001350) to more fully occupy the SSB-Ct binding site on DnaGC. Alternatively, optimisation of indole scaffold could provide an additional approach. Thermodynamic profiling of compounds using isothermal titration calorimetry will assist in future optimisations by quantifying the entropic and enthalpic contributions to binding. Eventually, i*n vivo* functional assays and estimates of antibacterial activities will be conducted on optimised analogues to asses their activity *in vivo*.

Finally, some of the compounds identified from the STD-NMR screen potentially bind to the RPD domain of DnaG primase. Additional experiments are needed in order to confirm that these compounds bind to RPD and that they do so in a manner that could inhibit RNA polymerase activity. Available crystal structure of *E. coli* and *S. aureus* species RPD domains indicate high structural similarity (Rymer *et al.*, 2012; Mansfield and Dixon, 2012). Therefore, potentially, these structures could be used for inhibitor design.

References

Aksimentiev, A., Schulten, K. (2005) Imaging α-Hemolysin with Molecular Dynamics: Ionic Conductance, Osmotic Permeability, and the Electrostatic Potential Map. *Biophysical journal*. **88**(6), 3745–3761.

Allegrozzi, M., Bertini, I., Janik, M.B.L., Lee, Y.M., Liu, G., Luchinat, C. (2000) Lanthanide-induced pseudocontact shifts for solution structure refinements of macromolecules in shells up to 40 Å from the metal ion. *Journal of the American Chemical Society*. **122**(17), 4154–4161.

Allen, F.H., Groom, C.R., Liebeschuetz, J.W., Bardwell, D.A., Olsson, T.S.G., Wood, P.A. (2012) The hydrogen bond environments of 1 H -tetrazole and tetrazolate rings: The structural basis for tetrazole-carboxylic acid bioisosterism. *Journal of Chemical Information and Modeling*. **52**(3), 857–866.

Angulo, J., Enríquez-Navas, P.M., Nieto, P.M. (2010) Ligand-receptor binding affinities from saturation transfer difference (STD) NMR spectroscopy: the binding isotherm of STD initial growth rates. *Chemistry (Weinheim an der Bergstrasse, Germany)*. **16**(26), 7803–7812.

Angulo, J., Nieto, P.M. (2011) STD-NMR: Application to transient interactions between biomolecules-a quantitative approach. *European Biophysics Journal*. **40**(12), 1357–1369.

Anon (2000) DnaB Helicase Affects the Initiation Specifity of Escherichia coli Primase on Single-Stranded DNA Templates†. **39**(4), 745–752.

Antony, E., Weiland, E.A., Korolev, S., Lohman, T.M. (2012) Plasmodium falciparum SSB tetramer wraps single-stranded DNA with similar topology but opposite polarity to E. coli SSB. *Journal of Molecular Biology*. **420**(4-5), 269–283.

Arad, G., Hendel, A., Urbanke, C., Curth, U., Livneh, Z. (2008) Single-stranded DNA-binding Protein Recruits DNA Polymerase V to Primer Termini on RecA-coated DNA. *Journal of Biological Chemistry*. **283**(13), 8274–8282.

Arai, K., Kornberg, A. (1979) A general priming system employing only dnaB protein and primase for DNA replication. *Proceedings of the National Academy of Sciences*. **76**(9), 4308–4312.

Aravind, L., Leipe, D.D., Koonin, E.V. (1998) Toprim--a conserved catalytic domain in type IA and II topoisomerases, DnaG-type primases, OLD family nucleases and RecR proteins. *Nucleic Acids Research*. **26**(18), 4205–4213.

Bailey, S., Eliason, W.K., Steitz, T.A. (2007) Structure of Hexameric DnaB Helicase and Its Complex with a Domain of DnaG Primase. *Science*. **318**(5849), 459–463.

Baker, T.A., Funnell, B.E., Kornberg, A. (1987) Helicase action of dnaB protein during replication from the Escherichia coli chromosomal origin in vitro. *Journal of Biological Chemistry*. **262**(14), 6877–6885.

Bárcena, M., Ruiz, T., Donate, L.E., Brown, S.E., Dixon, N., Radermacher, M., Carazo, J.M. (2001) The DnaB.DnaC complex: a structure based on dimers assembled around an occluded channel. *EMBO J*. **20**(6), 1462–1468.

Begley, D.W., Moen, S.O., Pierce, P.G., Zartler, E.R. (2013) Saturation transfer difference NMR for fragment screening. *Current protocols in chemical biology*. **5**(4), 251–268.

Berg, J.M. (1990) Zinc fingers and other metal-binding domains. Elements for interactions between macromolecules. *Journal of Biological Chemistry*. **265**(12), 6513–6516.

Bertini, I., Luchinat, C., Parigi, G. (2002) Magnetic susceptibility in paramagnetic NMR. *Progress in Nuclear Magnetic Resonance Spectroscopy*. **40**(3), 249–273.

Betz, M., Saxena, K., Schwalbe, H. (2006) Biomolecular NMR: a chaperone to drug discovery. *Current Opinion in Chemical Biology*. **10**(3), 219–225.

Bouché, J.P., Zechel, K., Kornberg, A. (1975) dnaG gene product, a rifampicin-resistant RNA polymerase, initiates the conversion of a single-stranded coliphage DNA to its duplex replicative form. *Journal of Biological Chemistry*. **250**(15), 5995–6001.

Buss, J.A., Kimura, Y., Bianco, P.R. (2008) RecG interacts directly with SSB: implications for stalled replication fork regression. *Nucleic Acids Research*. **36**(22), 7029–7042.

Bussi, G., Parrinello, M. (2007) Accurate sampling using Langevin dynamics. *Physical review. E, Statistical, nonlinear, and soft matter physics*. **75**(5 Pt 2), 056707.

Butland, G., Peregrin-Alvarez, J.M., Li, J., Yang, W., Yang, X., Canadien, V., Starostine, A., Richards, D., Beattie, B., Krogan, N., Davey, M., Parkinson, J., Greenblatt, J., Emili, A. (2005) Interaction network containing conserved and essential protein complexes in Escherichia coli. *Nature*. **433**(7025), 531–537.

Cadman, C.J., McGlynn, P. (2004) PriA helicase and SSB interact physically and functionally. *Nucleic Acids Research*. **32**(21), 6378–6387.

Chen, H., Viel, S., Ziarelli, F., Peng, L. (2013) 19F NMR: a valuable tool for studying biological events. *Chemical Society Reviews*. **42**(20), 7971.

Clore, G.M., Tang, C., Iwahara, J. (2007) Elucidating transient macromolecular interactions using paramagnetic relaxation enhancement. *Current Opinion in Structural Biology*. **17**(5), 603–616.

Cole, S.T. (2014) Who will develop new antibacterial agents? *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*. **369**(1645), 20130430–20130430.

Congreve, M., Carr, R., Murray, C., Jhoti, H. (2003) A 'rule of three' for fragment-based lead discovery? *Drug Discovery Today*. **8**(19), 876–877.

Cook, D., Brown, D., Alexander, R., March, R., Morgan, P., Satterthwaite, G., Pangalos, M.N. (2014) Lessons learned from the fate of AstraZeneca's drug pipeline: a five-dimensional framework. *Nat Rev Drug Discov*. **13**(6), 419–431.

Cooper, M.A. (2002) Optical biosensors in drug discovery. *Nat Rev Drug Discov*. **1**(7),

515–528.

Coskun-Ari, F.F., Hill, T.M. (1997) Sequence-specific Interactions in the Tus-Ter Complex and the Effect of Base Pair Substitutions on Arrest of DNA Replication in Escherichia coli. *Journal of Biological Chemistry*. **272**(42), 26448–26456.

Coskun-Ari, F.F., Skokotas, A., Moe, G.R., Hill, T.M. (1994) Biophysical characteristics of Tus, the replication arrest protein of Escherichia coli. *Journal of Biological Chemistry*. **269**(6), 4027–4034.

Dalvit, C. (2009) NMR methods in fragment screening: theory and a comparison with other biophysical techniques. *Drug Discovery Today*. **14**(21–22), 1051–1057.

Dalvit, C., Ardini, E., Fogliatto, G.P., Mongelli, N., Veronesi, M. (2004) Reliable high-throughput functional screening with 3-FABS. *Drug Discovery Today*. **9**(14), 595–602.

Dalvit, C., Fogliatto, G., Stewart, A., Veronesi, M., Stockman, B. (2001) WaterLOGSY as a method for primary NMR screening: practical aspects and range of applicability. *Journal of Biomolecular NMR*. **21**(4), 349–359.

Dalvit, C., Invernizzi, C., Vulpetti, A. (2014) Fluorine as a Hydrogen-Bond Acceptor: Experimental Evidence and Computational Calculations. *Chemistry-A European Journal*. **20**(35), 11058–11068.

Dalvit, C., Stockman, B.J. (2002) NMR screening techniques in drug discovery and drug design. *Progress in Nuclear Magnetic Resonance Spectroscopy*. **41**(3-4), 187–231.

Danielson, U.H. (2009) Fragment library screening and lead characterization using SPR biosensors. *Current Topics in Medicinal Chemistry*. **9**(18), 1725–1735.

Darden, T., York, D., Pedersen, L. (1993) Particle mesh Ewald: An N [center-dot] log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics*. **98**(12), 10089–10092.

Davey, N.E., Travé, G., Gibson, T.J. (2011) How viruses hijack cell regulation. *Trends in Biochemical Sciences*. **36**(3), 159–169.

Davis, A.M., Oprea, T.I., Teague, S.J., Leeson, P.D. (2001) Is There a Difference between Leads and Drugs? A Historical Perspective. *Journal of Chemical Information and Modeling*. **41**(5), 1308–1315.

Davis, B.J., Erlanson, D.A. (2013) Learning from our mistakes: The 'unknown knowns' in fragment screening. *Bioorganic and Medicinal Chemistry Letters*. **23**(10), 2844–2852.

Dias, D.M., Van Molle, I., Baud, M.G.J., Galdeano, C., Geraldes, C.F.G.C., Ciulli, A. (2014) Is NMR Fragment Screening Fine-Tuned to Assess Druggability of Protein-Protein Interactions? *ACS Medicinal Chemistry Letters*. **5**(1), 23–28.

Didenko, T., Liu, J.J., Horst, R., Stevens, R.C., Wüthrich, K. (2013) Fluorine-19 NMR of integral membrane proteins illustrated with studies of GPCRs. *Current Opinion in Structural Biology*. **23**(5), 740–747.

Doak, B.C., Morton, C.J., Simpson, J.S., Scanlon, M.J. (2013) Design and evaluation of the performance of an NMR screening fragment library. *Australian Journal of Chemistry*. **66**(12), 1465–1472.

Dolezal, O., Doughty, L., Hattarki, M.K., Fazio, V.J., Caradoc-Davies, T.T., Newman, J., Peat, T.S. (2013) Fragment screening for the modelling community: SPR, ITC, and crystallography. *Australian Journal of Chemistry*. **66**(12), 1507–1517.

Doyle, M. (1997) Characterization of binding interactions by isothermal titration calorimetry. *Current Opinion in Biotechnology*. **8**(1), 31–35.

Durrant, J.D., McCammon, J.A. (2011) Molecular dynamics simulations and drug discovery. *BMC Biology*. **9**(1), 71.

Duschinsky, R., Pleven, E., Heidelberger, C. (1957) The synthesis of 5-fluoropyrimidines. *Journal of the American Chemical Society*. **79**(16), 4559–4560.

Emsley, P., Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D*. **60**(12-1), 2126–2132.

Fang, L., Davey, M.J., O'Donnell, M. (1999) Replisome Assembly at oriC, the Replication Origin of E. coli, Reveals an Explanation for Initiation Sites outside an Origin. *Molecular Cell*. **4**(4), 541–553.

Feller, S.E., Zhang, Y., Pastor, R.W., Brooks, B.R. (1995) Constant pressure molecular dynamics simulation: The Langevin piston method. *The Journal of Chemical Physics*. **103**(11), 4613–4621.

Fernández, C., Wider, G. (2003) TROSY in NMR studies of the structure and function of large biological macromolecules. *Current Opinion in Structural Biology*. **13**(5), 570–580.

Fielding, L. (2007) NMR methods for the determination of protein-ligand dissociation constants. *Progress in Nuclear Magnetic Resonance Spectroscopy*. **51**(4), 219–242.

Fischbach, M.A., Walsh, C.T. (2009) Antibiotics for Emerging Pathogens. *Science*. **325**(5944), 1089–1093.

Furukohri, A., Nishikawa, Y., Tatsumi Akiyama, M., Maki, H. (2012) Interaction between Escherichia coli DNA polymerase IV and single-stranded DNA-binding protein is required for DNA synthesis on SSB-coated DNA. *Nucleic Acids Research*. **40**(13), 6039–6048.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M., Appel, R., Bairoch, A. (2005) Protein Identification and Analysis Tools on the ExPASy Server. In J. Walker, ed. *The Proteomics Protocols Handbook*. Humana Press, pp. 571–607.

George, N.P., Ngo, K.V., Chitteni-Pattu, S., Norais, C.A., Battista, J.R., Cox, M.M., Keck, J.L. (2012) Structure and Cellular Dynamics of Deinococcus radiodurans Single-stranded DNA (ssDNA)-binding Protein (SSB)-DNA Complexes. *Journal of Biological Chemistry*. **287**(26), 22123–22132.

Gribbon, Philip, Sewing, A. (2003) Fluorescence readouts in HTS: no gain without pain? *Drug Discovery Today*. **8**(22), 1035–1043.

Gribbon, Phillip, Andreas, S. (2005) High-throughput drug discovery: What can we expect from HTS? *Drug Discovery Today*. **10**(1), 17–22.

Grzesiek, S., Sass, H.-J. (2009) From biomolecular structure to functional understanding: new NMR developments narrow the gap. *Current Opinion in Structural Biology*. **19**(5), 585–595.

Guan, J.-Y., Keizers, P.H.J., Liu, W.-M., Löhr, F., Skinner, S.P., Heeneman, E.A., Schwalbe, H., Ubbink, M., Siegal, G. (2013) Small-Molecule Binding Sites on Proteins Established by Paramagnetic NMR Spectroscopy. *Journal of the American Chemical Society*. **135**(15), 5859–5868.

Hagmann, W.K. (2008) The Many Roles for Fluorine in Medicinal Chemistry. *Journal of Medicinal Chemistry*. **51**(15), 4359–4369.

Hajduk, P.J. (2006) SAR by NMR: Putting the Pieces Together. *Molecular interventions*. **6**(5), 266–272.

Hajduk, P.J., Boyd, S., Nettesheim, D., Nienaber, V., Severin, J., Smith, R., Davidson, D., Rockway, T., Fesik, S.W. (2000) Identification of Novel Inhibitors of Urokinase via NMR-Based Screening. *Journal of Medicinal Chemistry*. **43**(21), 3862–3866.

Hajduk, P.J., Olejniczak, E.T., Fesik, S.W. (1997) One-Dimensional Relaxation- and Diffusion-Edited NMR Methods for Screening Compounds That Bind to Macromolecules. *Journal of the American Chemical Society*. **119**(50), 12257–12261.

Han, E.S., Cooper, D.L., Persky, N.S., Sutera, V.A., Whitaker, R.D., Montello, M.L., Lovett, S.T. (2006) RecJ exonuclease: substrates, products and interaction with SSB. *Nucleic Acids Research*. **34**(4), 1084–1091.

Handa, P., Acharya, N., Varshney, U. (2001) Chimeras between Single-stranded DNA-binding Proteins from Escherichia coli and Mycobacterium tuberculosis Reveal That Their C-terminal Domains Interact with Uracil DNA Glycosylases. *Journal of Biological Chemistry*. **276**(20), 16992–16997.

Hass, M.A.S., Ubbink, M. (2014) Structure determination of protein-protein complexes with long-range anisotropic paramagnetic NMR restraints. *Current Opinion in Structural Biology*. **24**(1), 45–53.

Heinemann, J.A., Ankenbauer, R.G., Amábile-Cuevas, C.F. (2000) Do antibiotics maintain antibiotic resistance? *Drug Discovery Today*. **5**(5), 195–204.

Hodskinson, M.R.G., Allen, L.M., Thomson, D.P., Sayers, J.R. (2007) Molecular interactions of Escherichia coli ExoIX and identification of its associated 3′–5′ exonuclease activity. *Nucleic Acids Research*. **35**(12), 4094–4102.

Holdgate, G.A., Ward, W.H.J. (2005) Measurements of binding thermodynamics in drug discovery. *Drug Discovery Today*. **10**(22), 1543–1550.

Hopkins, A.L., Groom, C.R., Alex, A. (2004) Ligand efficiency: a useful metric for lead selection. *Drug Discovery Today*. **9**(10), 430–431.

Howard, S., Amin, N., Benowitz, A.B., Chiarparin, E., Cui, H., Deng, X., Heightman, T.D., Holmes, D.J., Hopkins, A., Huang, J., Jin, Q., Kreatsoulas, C., Martin, A.C.L., Massey, F., McCloskey, L., Mortenson, P.N., Pathuri, P., Tisi, D., Williams, P.A. (2013) Fragment-Based Discovery of 6-Azaindazoles As Inhibitors of Bacterial DNA Ligase. *ACS Medicinal Chemistry Letters*. **4**(12), 1208–1212.

Hubbard, R.E., Murray, J.B. (2011) *20 - Experiences in Fragment-Based Lead Discovery*. 1st ed. Elsevier Inc.

Humphrey, W., Dalke, A., Schulten, K. (1996) VMD: Visual molecular dynamics. *Journal of Molecular Graphics*. **14**(1), 33–38.

Irwin, J.J., Shoichet, B.K. (2005) ZINC − A Free Database of Commercially Available Compounds for Virtual Screening. *Journal of Chemical Information and Modeling*. **45**(1), 177–182.

Izaguirre, J.A., Catarello, D.P., Wozniak, J.M., Skeel, R.D. (2001) Langevin stabilization of molecular dynamics. *The Journal of Chemical Physics*. **114**(5), 2090–2098.

Jezewska, M.J., Rajendran, S., Bujalowska, D., Bujalowski, W. (1998) Does single-stranded DNA pass through the inner channel of the protein hexamer in the complex with the Escherichia coli DnaB Helicase? Fluorescence energy transfer studies. *Journal of Biological Chemistry*. **273**(17), 10515–10529.

John, M., Park, A.Y., Pintacuda, G., Dixon, N., Otting, G. (2005) Weak alignment of paramagnetic proteins warrants correction for residual CSA effects in measurements of pseudocontact shifts. *Journal of the American Chemical Society*. **127**(49), 17190–17191.

John, M., Pintacuda, G., Park, A.Y., Dixon, N. (2006) Structure determination of protein-ligand complexes by transferred paramagnetic shifts. *Journal of the American Chemical Society*. **128**(39), 12910–12916.

Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W., Klein, M.L. (1983) Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*. **79**(2), 926.

Kadaveru, K., Vyas, J., Schiller, M.R. (2008) Viral infection and human disease--insights from minimotifs. *Frontiers in bioscience : a journal and virtual library*. **13**, 6455–6471.

Keck, J.L., Berger, J.M. (2000) DNA replication at high resolution. *Chemistry and Biology*. **7**(3), R63–R71.

Keck, J.L., Roche, D.D., Lynch, A.S., Berger, J.M. (2000) Structure of the RNA Polymerase Domain of E. coli Primase. *Science*. **287**(5462), 2482–2486.

Keizers, P.H.J., Saragliadis, A., Hiruma, Y., Overhand, M., Ubbink, M. (2008) Design,

Synthesis, and Evaluation of a Lanthanide Chelating Protein Probe: CLaNP-5 Yields Predictable Paramagnetic Effects Independent of Environment. *Journal of the American Chemical Society*. **130**(44), 14802–14812.

Keizers, P.H.J., Ubbink, M. (2011) Paramagnetic tagging for protein structure and dynamics analysis. *Progress in Nuclear Magnetic Resonance Spectroscopy*. **58**(1-2), 88–96.

Kelly, A.E., Ou, H.D., Withers, R., Dötsch, V. (2002) Low-Conductivity Buffers for High-Sensitivity NMR Measurements. *Journal of the American Chemical Society*. **124**(40), 12013–12019.

Kelman, Z., Yuzhakov, A., Andjelkovic, J., O'Donnell, M. (1998) Devoted to the lagging strand-the [chi] subunit of DNA polymerase III holoenzyme contacts SSB to promote processive elongation and sliding clamp assembly. *EMBO J*. **17**(8), 2436–2449.

Klepeis, J.L., Lindorff-Larsen, K., Dror, R.O., Shaw, D.E. (2009) Long-timescale molecular dynamics simulations of protein structure and function. *Current Opinion in Structural Biology*. **19**(2), 120–127.

Koehler, J., Meiler, J. (2011) Expanding the utility of NMR restraints with paramagnetic compounds: Background and practical aspects. *Progress in Nuclear Magnetic Resonance Spectroscopy*. **59**(4), 360–389.

Kohara, Y., Tohdoh, N., Jiang, X.W. (1985) The distribution and properties of RNA primed initiation sites of DNA synthesis at the replication origin of Escherichia coli chromosome. *Nucleic Acids Research*.

Kozlov, A.G., Jezewska, M.J., Bujalowski, W., Lohman, T.M. (2010) Binding Specificity of Escherichia coli Single-Stranded DNA Binding Protein for the χ Subunit of DNA pol III Holoenzyme and PriA Helicase. *Biochemistry*. **49**(17), 3555–3566.

Krosky, D.J., Bianchet, M.A., Seiple, L., Chung, S., Amzel, L.M., Stivers, J.T. (2006) Mimicking damaged DNA with a small molecule inhibitor of human UNG2. *Nucleic Acids Research*. **34**(20), 5872–5879.

Krüger, D.M., Jessen, G., Gohlke, H. (2012) How good are state-of-the-art docking tools in predicting ligand binding modes in protein-protein interfaces? *Journal of Chemical Information and Modeling*. **52**(11), 2807–2811.

Kuntz, I.D., Chen, K., Sharp, K.A., Kollman, P.A. (1999) The maximal affinity of ligands. *Proceedings of the National Academy of Sciences*. **96**(18), 9997–10002.

Kusakabe, T., Richardson, C.C. (1996) The Role of the Zinc Motif in Sequence Recognition by DNA Primases. *Journal of Biological Chemistry*. **271**(32), 19563–19570.

Kwan, A.H., Mobli, M., Gooley, P.R., King, G.F., Mackay, J.P. (2011) Macromolecular NMR spectroscopy for the non-spectroscopist. *FEBS Journal*. **278**(5), 687–703.

LaPlante, S.R., Aubry, N., Bolger, G., Bonneau, P., Carson, R., Coulombe, R., Sturino,

C., Beaulieu, P.L. (2013) Monitoring Drug Self-Aggregation and Potential for Promiscuity in Off-Target In Vitro Pharmacology Screens by a Practical NMR Strategy. *Journal of Medicinal Chemistry*. **56**(17), 7073–7083.

LaPlante, S.R., Carson, R., Gillard, J., Aubry, N., Coulombe, R., Bordeleau, S., Bonneau, P., Little, M., O'Meara, J., Beaulieu, P.L. (2013) Compound Aggregation in Drug Discovery: Implementing a Practical NMR Assay for Medicinal Chemists. *Journal of Medicinal Chemistry*. **56**(12), 5142–5150.

Lecointe, F., Serena, C., Velten, M., Costes, A., McGovern, S., Meile, J.-C., Errington, J., Ehrlich, S.D., Noirot, P., Polard, P. (2007) Anticipating chromosomal replication fork arrest: SSB targets repair DNA helicases to active forks. *EMBO J*. **26**(19), 4239–4251.

Lewis, K. (2013) Platforms for antibiotic discovery. *Nat Rev Drug Discov*. **12**(5), 371–387.

Li, H., Robertson, A.D., Jensen, J.H. (2005) Very fast empirical prediction and rationalization of protein pKa values. *Proteins*. **61**(4), 704–721.

Lian, L.-Y. (2013) NMR studies of weak protein-protein interactions. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 1–14.

Liepinsh, E., Otting, G. (1997) Organic solvents identify specific ligand binding sites on protein surfaces. *Nature Biotechnology*. **15**(3), 264–268.

Longley, D.B., Johnston, P.G. (2005) Molecular mechanisms of drug resistance. *The Journal of Pathology*. **205**(2), 275–292.

Loscha, K., Oakley, A.J., Bancia, B., Schaeffer, P.M., Prosselkov, P., Otting, G., Wilce, M.C.J., Dixon, N. (2004) Expression, purification, crystallization, and NMR studies of the helicase interaction domain of Escherichia coli DnaG primase. *Protein Expression and Purification*. **33**(2), 304–310.

Low, R.L., Shlomai, J., Kornberg, A. (1982) Protein n, a primosomal DNA replication protein of Escherichia coli. Purification and characterization. *Journal of Biological Chemistry*. **257**(11), 6242–6250.

Lu, D., Bernstein, D.A., Satyshur, K.A., Keck, J.L. (2010) Small-molecule tools for dissecting the roles of SSB/protein interactions in genome maintenance. *Proceedings of the National Academy of Sciences*. **107**(2), 633–638.

Lu, D., Keck, J.L. (2008) Structural basis of Escherichia coli single-stranded DNA-binding protein stimulation of exonuclease I. *Proceedings of the National Academy of Sciences*. **105**(27), 9169–9174.

Lu, D., Windsor, M.A., Gellman, S.H., Keck, J.L. (2009) Peptide Inhibitors Identify Roles for SSB C-Terminal Residues in SSB/Exonuclease I Complex Formation. *Biochemistry*. **48**(29), 6764–6771.

Luzuriaga, J. (2011) *The Conservation of χ and Single-Stranded DNA Binding Protein Interactions Between Escherichia coli and Acinetobacter baumannii*. University of

Wollongong.

Macarron, R., Banks, M.N., Bojanic, D., Burns, D.J., Cirovic, D.A., Garyantes, T., Green, D.V.S., Hertzberg, R.P., Janzen, W.P., Paslay, J.W., Schopfer, U., Sittampalam, G.S. (2011) Impact of high-throughput screening in biomedical research. *Nat Rev Drug Discov*. **10**(3), 188–195.

MacKerell, A.D., Bashford, D., Bellott, Dunbrack, R.L., Evanseck, J.D., Field, M.J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F.T.K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D.T., Prodhom, B., Reiher, W.E., Roux, B., Schlenkrich, M., Smith, J.C., Stote, R., Straub, J., Watanabe, M., Wiórkiewicz-Kuczera, J., Yin, D., Karplus, M. (1998) All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *The Journal of Physical Chemistry B*. **102**(18), 3586–3616.

Mackerell, A.D., Feig, M., Brooks, C.L. (2004) Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *Journal of Computational Chemistry*. **25**(11), 1400–1415.

Malik, M.A., Wani, M.Y., Al-Thabaiti, S.A., Shiekh, R.A. (2013) Tetrazoles as carboxylic acid isosteres: chemistry and biology. *Journal of Inclusion Phenomena and Macrocyclic Chemistry*. **78**(1-4), 15–37.

Mansfield, R.E., Dixon, N. (2012) Priming the Engine of DNA Synthesis. *Structure*. **20**(9), 1447–1448.

Marceau, A.H., Bahng, S., Massoni, S.C., George, N.P., Sandler, S.J., Marians, K.J., Keck, J.L. (2011) Structure of the SSB-DNA polymerase III interface and its role in DNA replication. *EMBO J*. **30**(20), 4236–4247.

Marceau, A.H., Bernstein, D.A., Walsh, B.W., Shapiro, W., Simmons, L.A., Keck, J.L. (2013) Protein Interactions in Genome Maintenance as Novel Antibacterial Targets S. Korolev, ed. *PLoS ONE*. **8**(3), e58765.

Mashalidis, E.H., Śledź, P., Lang, S., Abell, C. (2013) A three-stage biophysical screening cascade for fragment-based drug discovery. *Nature Protocols*. **8**(11), 2309–2324.

Mayer, M., James, T.L. (2002) Detecting ligand binding to a small RNA target via saturation transfer difference NMR experiments in D_2O and H_2O. *Journal of the American Chemical Society*. **124**(45), 13376–13377.

Mayer, M., Meyer, B. (1999) Characterization of Ligand Binding by Saturation Transfer Difference NMR Spectroscopy. *Angewandte Chemie International Edition*. **38**(12), 1784–1788.

Mayer, M., Meyer, B. (2001) Group epitope mapping by saturation transfer difference NMR to identify segments of a ligand in direct contact with a protein receptor. *Journal of the American Chemical Society*. **123**(25), 6108–6117.

McHenry, C.S. (2011) Bacterial replicases and related polymerases. *Current Opinion in*

Chemical Biology. **15**(5), 587–594.

McHenry, C.S., Crow, W. (1979) DNA polymerase III of Escherichia coli. Purification and identification of subunits. *Journal of Biological Chemistry*. **254**(5), 1748–1753.

McMacken, R., Ueda, K., Kornberg, A. (1977) Migration of Escherichia coli dnaB protein on the template DNA strand as a mechanism in initiating DNA replication. *Proceedings of the National Academy of Sciences*. **74**(10), 4190–4194.

Medek, A., Hajduk, P.J., Mack, J., Fesik, S.W. (2000) The Use of Differential Chemical Shifts for Determining the Binding Site Location and Orientation of Protein-Bound Ligands. *Journal of the American Chemical Society*. **122**(6), 1241–1242.

Meinecke, R., Meyer, B. (2001) Determination of the Binding Specificity of an Integral Membrane Protein by Saturation Transfer Difference NMR: RGD Peptide Ligands Binding to Integrin α IIbβ 3†. *Journal of Medicinal Chemistry*. **44**(19), 3059–3065.

Meyer, B., Peters, T. (2003) NMR Spectroscopy Techniques for Screening and Identifying Ligand Binding to Protein Receptors. *Angewandte Chemie International Edition*. **42**(8), 864–890.

Molineux, I.J., Gefter, M.L. (1974) Properties of the Escherichia coli DNA Binding (Unwinding) Protein: Interaction with DNA Polymerase and DNA. *Proceedings of the National Academy of Sciences*. **71**(10), 3858–3862.

Mooij, W.T.M., Hartshorn, M.J., Tickle, I.J., Sharff, A.J., Verdonk, M.L., Jhoti, H. (2006) Automated protein-ligand crystallography for structure-based drug design. *ChemMedChem*. **1**(8), 827–838.

Moore, R.L., Brubaker, R.R. (1975) Hybridization of Deoxyribonucleotide Sequences of Yersinia enterocolitica and Other Selected Members of Enterobacteriaceae. *International Journal of Systematic Bacteriology*. **25**(4), 336–339.

Morphy, R. (2010) Selectively Nonselective Kinase Inhibition: Striking the Right Balance. *Journal of Medicinal Chemistry*. **53**(4), 1413–1437.

Naue, N., Beerbaum, M., Bogutzki, A., Schmieder, P., Curth, U. (2013) The helicase-binding domain of Escherichia coli DnaG primase interacts with the highly conserved C-terminal region of single-stranded DNA-binding protein. *Nucleic Acids Research*. **41**(8), 4507–4517.

Naue, N., Fedorov, R., Pich, A., Manstein, D.J., Curth, U. (2011) Site-directed mutagenesis of the χ subunit of DNA polymerase III and single-stranded DNA-binding protein of E. coli reveals key residues for their interaction. *Nucleic Acids Research*. **39**(4), 1398–1407.

Neumann, T., Junker, H.-D., Schmidt, K., Sekul, R. (2007) SPR-based fragment screening: advantages and applications. *Current Topics in Medicinal Chemistry*. **7**(16), 1630–1642.

Newman, J., Egan, D., Walter, T.S., Meged, R. (2005) Towards rationalization of crystallization screening for small-to medium-sized academic laboratories: the

PACT/JCSG+ strategy. *... Section D: Biological ...*.

O'Donnell, M., Studwell, P.S. (1990) Total reconstitution of DNA polymerase III holoenzyme reveals dual accessory protein clamps. *Journal of Biological Chemistry*. **265**(2), 1179–1187.

Oakley, A.J., Loscha, K.V., Schaeffer, P.M., Liepinsh, E., Pintacuda, G., Wilce, M.C.J., Dixon, N. (2005) Crystal and solution structures of the helicase-binding domain of Escherichia coli primase. *Journal of Biological Chemistry*. **280**(12), 11495–11504.

Okazaki, R., Okazaki, T., Sakabe, K., Sugimoto, K. (1967) Mechanism of DNA replication possible discontinuity of DNA chain growth. *Japanese journal of medical science & biology*. **20**(3), 255–260.

Oprea, T.I. (2002) Chemical space navigation in lead discovery. *Current Opinion in Chemical Biology*. **6**(3), 384–389.

Otting, G. (2008) Prospects for lanthanides in structural biology by NMR. *Journal of Biomolecular NMR*. **42**(1), 1–9.

Otting, G. (2010) Protein NMR using paramagnetic ions. *Annual Review of Biophysics*. **39**(1), 387–405.

Pan, H., Wigley, D.B. (2000) Structure of the zinc-binding domain of Bacillus stearothermophilus DNA primase. *Structure*. **8**(3), 231–239.

Pardridge, W.M. (1995) Transport of small molecules through the blood-brain barrier: biology and methodology. *Advanced Drug Delivery Reviews*. **15**(1-3), 5–36.

Patching, S.G. (2014) Surface plasmon resonance spectroscopy for characterisation of membrane protein-ligand interactions and its potential for drug discovery. *Biochimica et Biophysica Acta (BBA) - Biomembranes*. **1838**(1 PARTA), 43–55.

Pellecchia, M. (2005) Solution Nuclear Magnetic Resonance Spectroscopy Techniques for Probing Intermolecular Interactions. *Chemistry and Biology*. **12**(9), 961–971.

Pellecchia, M., Sem, D.S., Wüthrich, K. (2002) NMR in drug discovery. *Nat Rev Drug Discov*. **1**(3), 211–219.

Petros, A.M., Swann, S.L., Song, D., Swinger, K., Park, C., Zhang, H., Wendt, M.D., Kunzer, A.R., Souers, A.J., Sun, C. (2014) Fragment-based discovery of potent inhibitors of the anti-apoptotic MCL-1 protein. *Bioorganic and Medicinal Chemistry Letters*. **24**(6), 1484–1488.

Phillips, J.C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R.D., Kalé, L., Schulten, K. (2005) Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry*. **26**(16), 1781–1802.

Pintacuda, G., John, M., Su, X.-C., Otting, G. (2007) NMR Structure Determination of Protein−Ligand Complexes by Lanthanide Labeling. *Accounts of Chemical Research*. **40**(3), 206–212.

Podobnik, M., McInerney, P., O'Donnell, M., Kuriyan, J. (2000) A TOPRIM domain in the crystal structure of the catalytic core of Escherichia coli primase confirms a structural link to DNA topoisomerases. *Journal of Molecular Biology*. **300**(2), 353–362.

Powers, J.H. (2004) Antimicrobial drug development – the past, the present, and the future. *Clinical Microbiology and Infection*. **10**, 23–31.

Purnapatre, K., Handa, P., Venkatesh, J., Varshney, U. (1999) Differential effects of single-stranded DNA binding proteins (SSBs) on uracil DNA glycosylases (UDGs) from Escherichia coli and mycobacteria. *Nucleic Acids Research*. **27**(17), 3487–3492.

Purser, S., Moore, P.R., Swallow, S., Gouverneur, V. (2008) Fluorine in medicinal chemistry. *Chemical Society Reviews*. **37**(2), 320.

Raghunathan, S., Ricard, C.S., Lohman, T.M., Waksman, G. (1997) Crystal structure of the homo-tetrameric DNA binding domain of Escherichia coli single-stranded DNA-binding protein determined by multiwavelength x-ray diffraction on the selenomethionyl protein at 2.9-A resolution. *Proceedings of the National Academy of Sciences*. **94**(13), 6652–6657.

Rees, D.C., Congreve, M., Murray, C.W., Carr, R. (2004) Fragment-based lead discovery. *Nat Rev Drug Discov*. **3**(8), 660–672.

Riek, R., Pervushin, K., Wüthrich, K. (2000) TROSY and CRINEPT: NMR with large molecular and supramolecular structures in solution. *Trends in Biochemical Sciences*. **25**(10), 462–468.

Robinson, A., Brzoska, A.J., Turner, K.M., Withers, R., Harry, E.J., Lewis, P.J., Dixon, N. (2010) Essential Biological Processes of an Emerging Pathogen: DNA Replication, Transcription, and Cell Division in Acinetobacter spp. *Microbiology and Molecular Biology Reviews*. **74**(2), 273–297.

Robinson, A., Dixon, N., Causer, R.J. (2012) Architecture and conservation of the bacterial DNA replication machinery, an underexploited drug target. *Current Drug Targets*. **13**(3), 352–372.

Robinson, A., van Oijen, A.M. (2013) Bacterial replication, transcription and translation: mechanistic insights from single-molecule biochemical studies. *Nature Reviews Microbiology*. **11**(5), 303–315.

Rymer, R.U., Solorio, F.A., Tehranchi, A.K., Chu, C., Corn, J.E., Keck, J.L., Wang, J.D., Berger, J.M. (2012) Binding Mechanism of Metal·NTP Substrates and Stringent-Response Alarmones to Bacterial DnaG-Type Primases. *Structure*. **20**(9), 1478–1489.

Ryzhikov, M., Koroleva, O., Postnov, D., Tran, A., Korolev, S. (2011) Mechanism of RecO recruitment to DNA by single-stranded DNA binding protein. *Nucleic Acids Research*. **39**(14), 6305–6314.

Sanyal, G., Doig, P. (2012) Bacterial DNA replication enzymes as targets for antibacterial drug discovery. *Expert Opinion on Drug Discovery*. **7**(4), 327–339.

Scaringi, C., Minniti, G., Caporello, P., Enrici, R.M. (2012) Integrin inhibitor

cilengitide for the treatment of glioblastoma: a brief overview of current clinical results. *Anticancer research*. **32**(10), 4213–4223.

Schaeffer, P.M., Headlam, M.J., Dixon, N. (2005) Protein - Protein Interactions in the Eubacterial Replisome. *IUBMB Life*. **57**(1), 5–12.

Schanda, P., Kupče, Ē., Brutscher, B. (2005) SOFAST-HMQC Experiments for Recording Two-dimensional Deteronuclear Correlation Spectra of Proteins within a Few Seconds. *Journal of Biomolecular NMR*. **33**(4), 199–211.

Schmitz, C., Stanton-Cook, M.J., Su, X.-C., Otting, G., Huber, T. (2008) Numbat: an interactive software tool for fitting Δχ-tensors to molecular coordinates using pseudocontact shifts. *Journal of Biomolecular NMR*. **41**(3), 179–189.

Scott, D.E., Coyne, A.G., Hudson, S.A., Abell, C. (2012) Fragment-based approaches in drug discovery and chemical biology. *Biochemistry*. **51**(25), 4990–5003.

Secchiero, P., Bosco, R., Celeghini, C., Zauli, G. (2011) Recent advances in the therapeutic perspectives of Nutlin-3. *Current pharmaceutical design*. **17**(6), 569–577.

Shaw, D.E., Maragakis, P., Lindorff-Larsen, K., Piana, S., Dror, R.O., Eastwood, M.P., Bank, J.A., Jumper, J.M., Salmon, J.K., Shan, Y., Wriggers, W. (2010) Atomic-Level Characterization of the Structural Dynamics of Proteins. *Science*. **330**(6002), 341–346.

Shereda, R.D., Bernstein, D.A., Keck, J.L. (2007) A Central Role for SSB in Escherichia coli RecQ DNA Helicase Function. *Journal of Biological Chemistry*. **282**(26), 19247–19258.

Shereda, R.D., Kozlov, A.G., Lohman, T.M., Cox, M.M., Keck, J.L. (2008) SSB as an organizer/mobilizer of genome maintenance complexes. *Critical Reviews in Biochemistry and Molecular Biology*. **43**(5), 289–318.

Shereda, R.D., Reiter, N.J., Butcher, S.E., Keck, J.L. (2009) Identification of the SSB Binding Site on E. coli RecQ Reveals a Conserved Surface for Binding SSB's C Terminus. *Journal of Molecular Biology*. **386**(3), 612–625.

Shishmarev, D., Otting, G. (2013) How reliable are pseudocontact shifts induced in proteins and ligands by mobile paramagnetic metal tags? A modelling study. *Journal of Biomolecular NMR*. **56**(3), 203–216.

Shishmarev, D., Wang, Y., Mason, C.E., Su, X.C., Oakley, A.J., Graham, B., Huber, T., Dixon, N., Otting, G. (2014) Intramolecular binding mode of the C-terminus of Escherichia coli single-stranded DNA binding protein determined by nuclear magnetic resonance spectroscopy. *Nucleic Acids Research*. **42**(4), 2750–2757.

Stark, J., Powers, R. (2008) Rapid protein-ligand costructures using chemical shift perturbations. *Journal of the American Chemical Society*. **130**(2), 535–545.

Stukenberg, P.T., O'Donnell, M. (1995) Assembly of a Chromosomal Replication Machine: Two DNA Polymerases, a Clamp Loader, and Sliding Clamps in One Holoenzyme Particle. *Journal of Biological Chemistry*. **270**(22), 13384–13391.

Su, X.-C., Liang, H., Loscha, K.V., Otting, G. (2009) [Ln(DPA) 3] 3−Is a Convenient Paramagnetic Shift Reagent for Protein NMR Studies. *Journal of the American Chemical Society*. **131**(30), 10352–10353.

Su, X.-C., Schaeffer, P.M., Loscha, K.V., Gan, P.H.P., Dixon, N., Otting, G. (2006) Monomeric solution structure of the helicase-binding domain of Escherichia coli DnaG primase. *FEBS Journal*. **273**(21), 4997–5009.

Su, X.-C., Wang, Y., Yagi, H., Shishmarev, D., Mason, C.E., Smith, P.J., Vandevenne, M., Dixon, N., Otting, G. (2014) Bound or Free: Interaction of the C-Terminal Domain of Escherichia coliSingle-Stranded DNA-Binding Protein (SSB) with the Tetrameric Core of SSB. *Biochemistry*. **53**(12), 1925–1934.

Tak Lo, A. (2012) *Protein dynamics on the lagging strand during DNA synthesis*. Australia: University of Wollongong.

Trott, O., Olson, A.J. (2010) Software news and update AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry*. **31**(2), 455–461.

Umezu, K., Kolodner, R.D. (1994) Protein interactions in genetic recombination in Escherichia coli. Interactions involving RecO and RecR overcome the inhibition of RecA by single-stranded DNA-binding protein. *Journal of Biological Chemistry*. **269**(47), 30005–30013.

Veber, D.F., Johnson, S.R., Cheng, H.-Y., Smith, B.R., Ward, K.W., Kopple, K.D. (2002) Molecular properties that influence the oral bioavailability of drug candidates. *Journal of Medicinal Chemistry*. **45**(12), 2615–2623.

Verdonk, M.L., Rees, D.C. (2008) Group efficiency: A guideline for hits-to-leads chemistry. *ChemMedChem*. **3**(8), 1179–1180.

Via, A., Uyar, B., Brun, C., Zanzoni, A. (2015) How pathogens use linear motifs to perturb host cell networks. *Trends in Biochemical Sciences*. **40**(1), 36–48.

Viegas, A., Manso, J., Nobrega, F.L., Cabrita, E.J. (2011) Saturation-Transfer Difference (STD) NMR: A Simple and Fast Method for Ligand Screening and Characterization of Protein Binding. *Journal of Chemical Education*. **88**(7), 990–994.

Volkov, A.N., Worrall, J. (2006) Solution structure and dynamics of the complex between cytochrome c and cytochrome c peroxidase determined by paramagnetic NMR. *Proceedings of the National Academy of Sciences*.

Vom, A., Headey, S., Wang, G., Capuano, B., Yuriev, E., Scanlon, M.J., Simpson, J.S. (2013) Detection and Prevention of Aggregation-based False Positives in STD-NMR-based Fragment Screening. *Australian Journal of Chemistry*. **66**(12), 1518.

Vranken, W.F., Boucher, W., Stevens, T.J., Fogh, R.H., Pajon, A., Llinas, M., Ulrich, E.L., Markley, J.L., Ionides, J., Laue, E.D. (2005) The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins*. **59**(4), 687–696.

Vulpetti, A., Hommel, U., Landrum, G., Lewis, R., Dalvit, C. (2009) Design and NMR-

Based Screening of LEF, a Library of Chemical Fragments with Different Local Environment of Fluorine. *Journal of the American Chemical Society*. **131**(36), 12949–12959.

Wagstaff, J.L., Vallath, S., Marshall, J.F., Williamson, R.A., Howard, M.J. (2010) Two-dimensional heteronuclear saturation transfer difference NMR reveals detailed integrin αvβ6 protein-peptide interactions. *Chemical Communications*. **46**(40), 7533–7535.

Walters, W.P., Namchuk, M. (2003) Designing screens: how to make your hits a hit. *Nat Rev Drug Discov*. **2**(4), 259–266.

Wang, B., Westerhoff, L.M., Merz, K.M. (2007) A Critical Assessment of the Performance of Protein−Ligand Scoring Functions Based on NMR Chemical Shift Perturbations. *Journal of Medicinal Chemistry*. **50**(21), 5128–5134.

Warren, G.L., Do, T.D., Kelley, B.P., Nicholls, A., Warren, S.D. (2012) Essential considerations for using protein-ìligand structures in drug discovery. *Drug Discovery Today*. **17**(23,Äì24), 1270–1281.

Wells, J.A., McClendon, C.L. (2007) Reaching for high-hanging fruit in drug discovery at protein–protein interfaces. *Nature*. **450**(7172), 1001–1009.

Williams, N.K. (2002) In Vivo Protein Cyclization Promoted by a Circularly Permuted Synechocystis sp. PCC6803 DnaB Mini-intein. *Journal of Biological Chemistry*. **277**(10), 7790–7798.

Williamson, M.P. (2013) Using chemical shift perturbation to characterise ligand binding. *Progress in Nuclear Magnetic Resonance Spectroscopy*. **73**, 1–16.

Witte, G., Urbanke, C., Curth, U. (2003) DNA polymerase III χ subunit ties single-stranded DNA binding protein to the bacterial replication machinery. *Nucleic Acids Research*. **31**(15), 4434–4440.

Yagi, H., Pilla, K.B., Maleckis, A., Graham, B., Huber, T., Otting, G. (2013) Three-dimensional protein fold determination from backbone amide pseudocontact shifts generated by lanthanide tags at multiple sites. *Structure*. **21**(6), 883–890.

Yang, C.-W. (2012) A comparative study of short linear motif compositions of the influenza A virus ribonucleoproteins. E. G. Brown, ed. *PLoS ONE*. **7**(6), e38637.

Yuzhakov, A., Kelman, Z., O'Donnell, M. (1999) Trading Places on DNA—A Three-Point Switch Underlies Primer Handoff from Primase to the Replicative DNA Polymerase. *Cell*. **96**(1), 153–163.

Yuzhakov, A., Turner, J., O'Donnell, M. (1996) Replisome Assembly Reveals the Basis for Asymmetric Function in Leading and Lagging Strand Replication. *Cell*. **86**(6), 877–886.

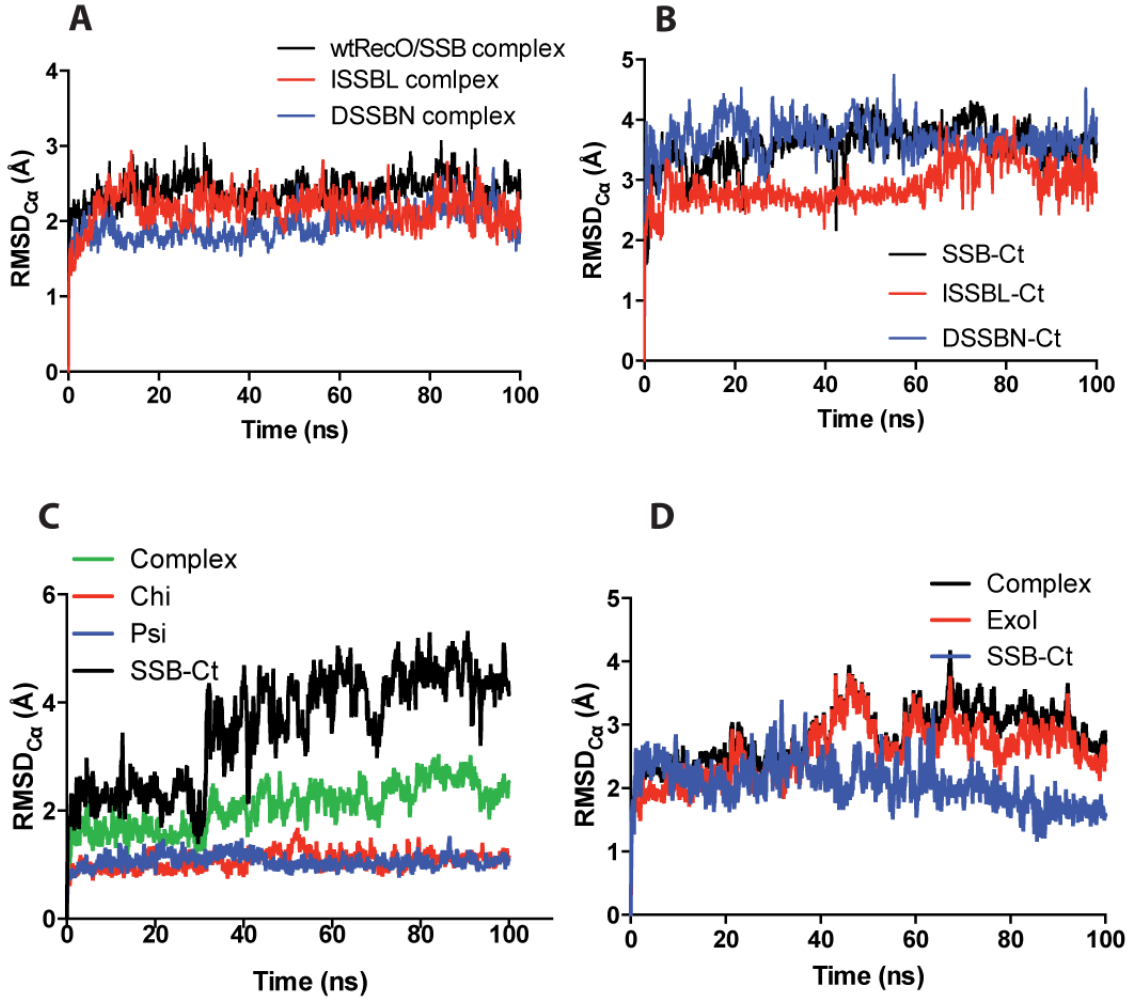
Zhu, G., Yao, X. (2008) TROSY-based NMR experiments for NMR studies of large biomolecules. *Progress in Nuclear Magnetic Resonance Spectroscopy*. **52**(1), 49–68.

Zuiderweg, E.R.P. (2002) Mapping Protein−Protein Interactions in Solution by NMR

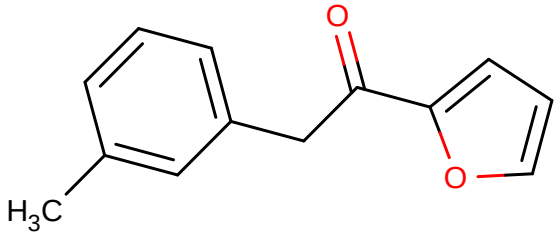
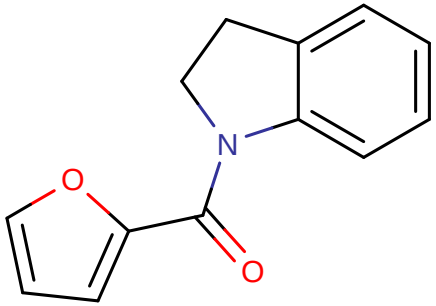
Spectroscopy. *Biochemistry*. **41**(1), 1–7.

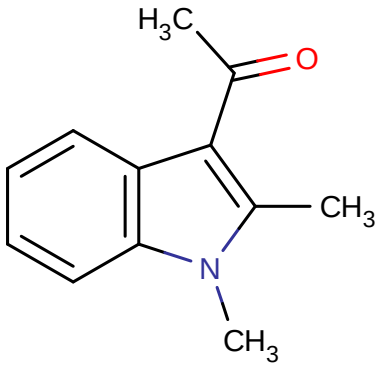
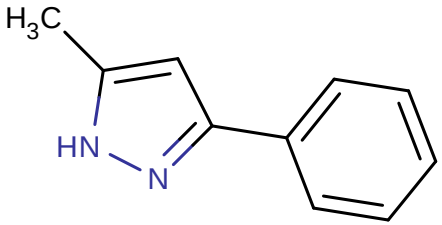
Appendices

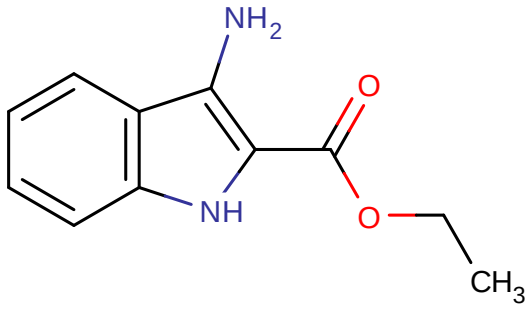
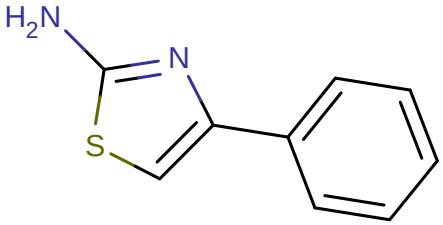
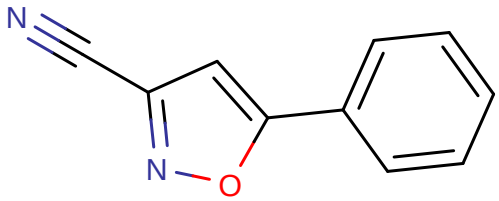
Appendix A. *Backbone RMSD plotted for each individual chain of RecO/SSB-Ct, χ/SSB-Ct and ExoI/SSB-Ct systems along the 100 ns MD simulations. (A) represents RMSDs for the entire RecO/wtSSB-Ct complex (black), RecO/DDDLPF (red), and RecO/NNNIPF (blue) mutated systems, (B) is RMSD for the SSB-Ct peptides from RecO/SSB-Ct complex. (C) represents χ/ψ-SSB-Ct complex, where the complex is in green, χ in red, psi in blue and SSB-Ct in black, (D) Black is ExoI/SSB-Ct complex (black), red is ExoI and blue SSB-Ct respectively.*

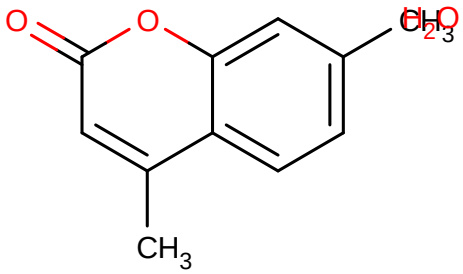
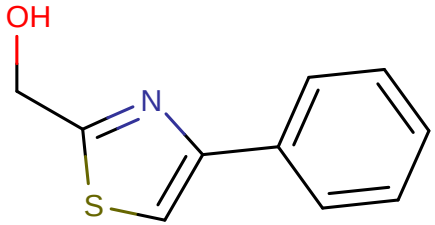
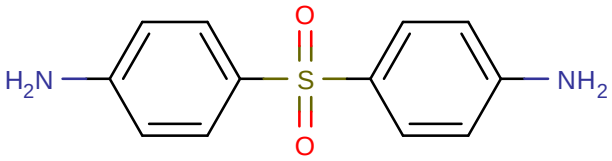


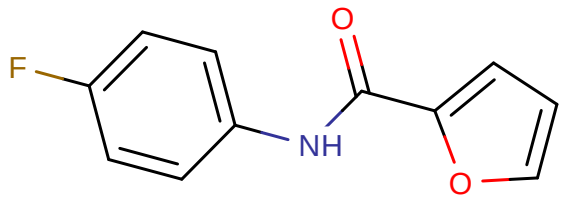
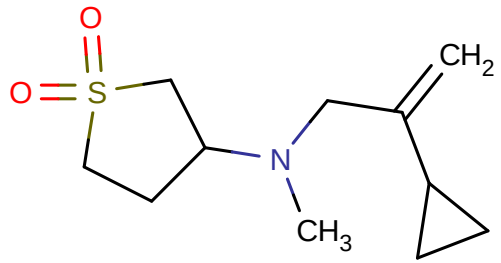
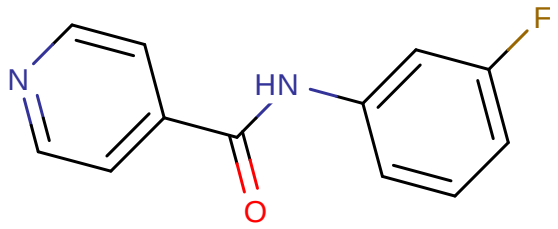
Appendix B. *List of compounds identified as STD rank 2 and rank 3.*
STD rank 2

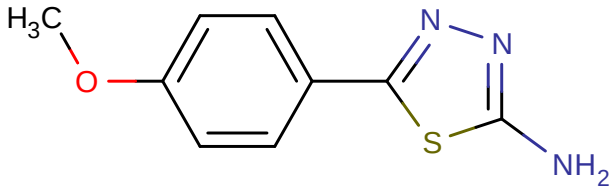
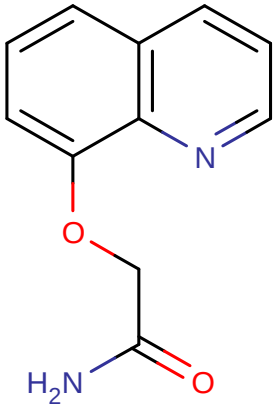
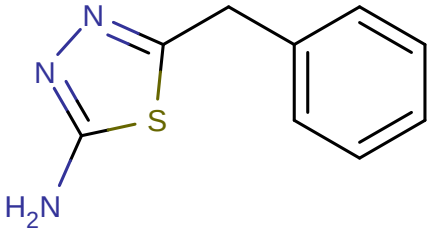
Structure	Plate Name	Plate Well	Mol Weight	Formula	IUPAC Name	LogP	Rotatable bonds	Molecule Name
	2011_ Chembridge 1	A10	201.22	C12H11NO2	N-(3-methylphenyl)furan-2-carboxamide	2.64	2	MIPS-0000369
	2011_ Chembridge 1	C7	213.23	C13H11NO2	1-(furan-2-carbonyl)-2,3-dihydro-1H-indole	2.03	1	MIPS-0000386

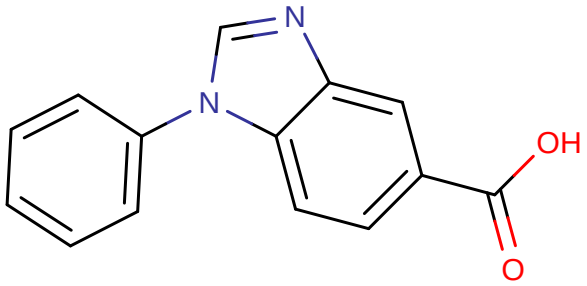
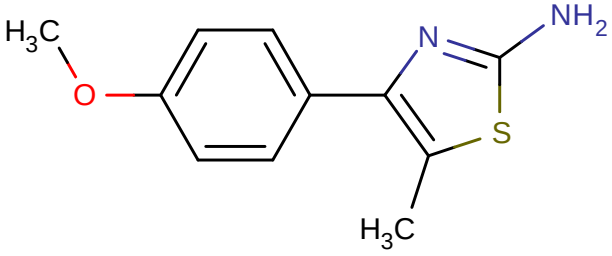
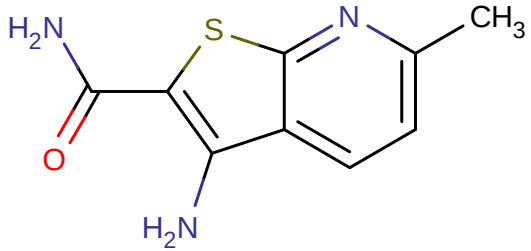
Structure	Plate Name	Plate Well	Mol Weight	Formula	IUPAC Name	LogP	Rotatable bonds	Molecule Name
	2011_ Chembridge 1	B6	187.24	C12H13NO	1-(1,2-dimethyl-1H-indol-3-yl)ethan-1-one	2.05	1	MIPS-0000387
	2011_ Chembridge 1	B12	158.20	C10H10N2	5-methyl-3-phenyl-1H-pyrazole	2.51	1	MIPS-0000392

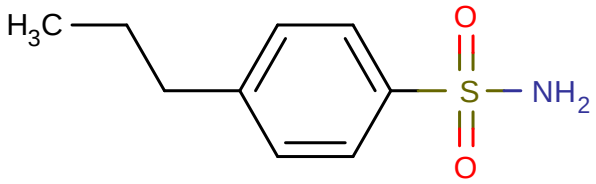
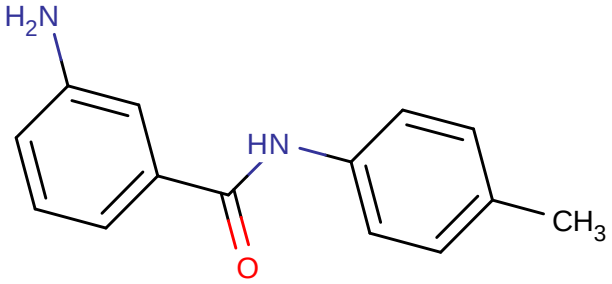
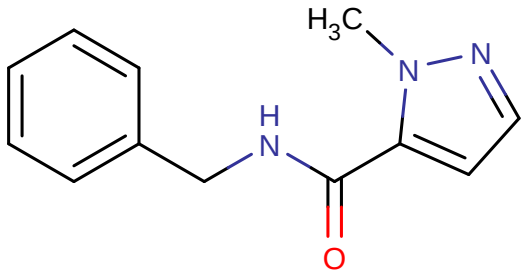
Structure	Plate Name	Plate Well	Mol Weight	Formula	IUPAC Name	LogP	Rotatable bonds	Molecule Name
	2011_Life3	C12	204.23	C11H12N2O2	ethyl 3-amino-1H-indole-2-carboxylate	2.17	3	MIPS-0000417
	2011_Chembridge 1	D9	176.24	C9H8N2S	4-phenyl-1,3-thiazol-2-amine	2.52	1	MIPS-0000424
	2011_Life3	D8	184.19	C11H8N2O	2-(5-phenyl-1,2-oxazol-3-yl)acetonitrile	1.99	2	MIPS-0000428

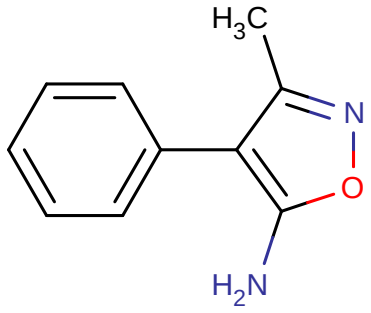
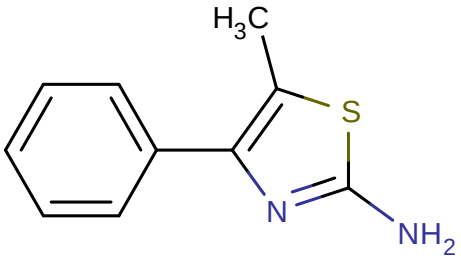
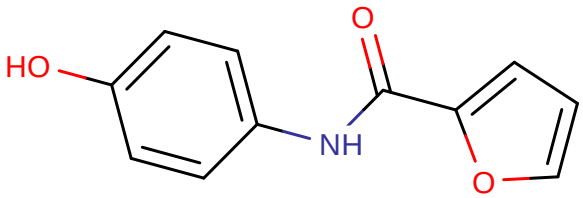
Structure	Plate Name	Plate Well	Mol Weight	Formula	IUPAC Name	LogP	Rotatable bonds	Molecule Name
	2011_ Life3	F1	176.17	C10H8O3	7-hydroxy-4-methyl-2H-chromen-2-one	1.78	0	MIPS-0000445
	2011_ Chembridge 1	F12	191.25	C10H9NOS	(4-phenyl-1,3-thiazol-2-yl)methanol	1.97	2	MIPS-0000463
	2011_ Chembridge 1	H9	248.30	C12H12N2O2S	4-(4-aminobenzenesulfonyl)aniline	1.27	2	MIPS-0000487

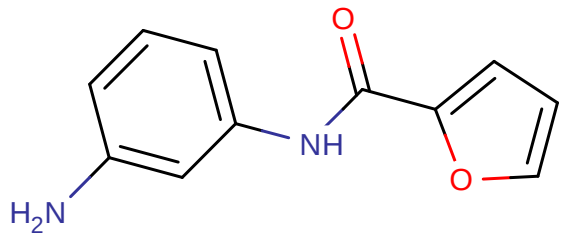
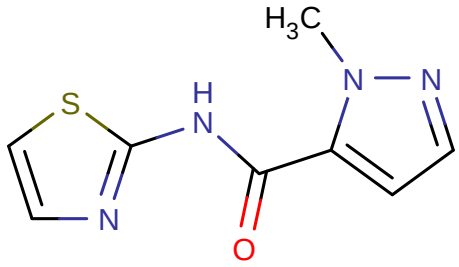
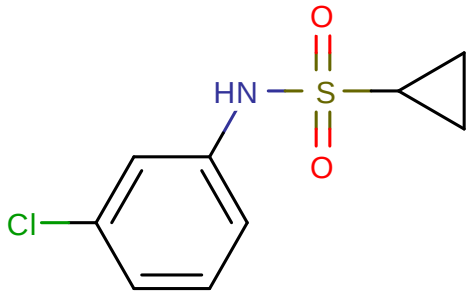
Structure	Plate Name	Plate Well	Mol Weight	Formula	IUPAC Name	LogP	Rotatable bonds	Molecule Name
	2011_ Chembridge 2	B7	205.19	C11H8FNO2	N-(4-fluorophenyl)furan-2-carboxamide	2.27	2	MIPS-0000520
	2011_ Life4	C11	217.29	C9H15NO3S	N-(1,1-dioxo-1λ⁶ -thiolan-3-yl)-N-methylcyclopropanecarboxamide	-1.02	2	MIPS-0000532
	2011_ Chembridge 2	C7	216.21	C12H9FN2O	N-(3-fluorophenyl)pyridine-4-carboxamide	1.99	2	MIPS-0000544

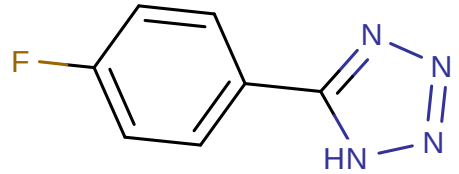
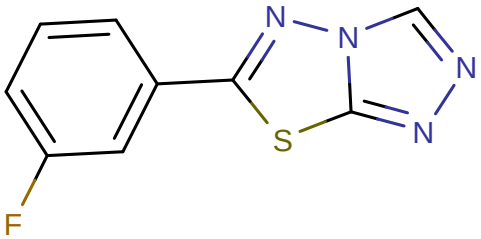
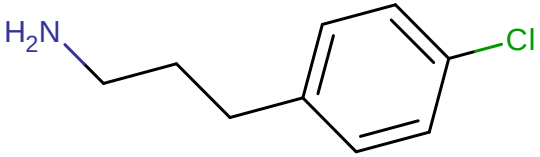
Structure	Plate Name	Plate Well	Mol Weight	Formula	IUPAC Name	LogP	Rotatable bonds	Molecule Name
	2011_ Chembridge 2	D9	207.25	C9H9N3OS	5-(4-methoxyphenyl)-1,3,4-thiadiazol-2-amine	1.47	2	MIPS-0000556
	2011_ Chembridge 2	D4	202.21	C11H10N2O2	2-(quinolin-8-yloxy)acetamide	0.64	3	MIPS-0000564
	2011_ Chembridge 2	E4	191.25	C9H9N3S	5-benzyl-1,3,4-thiadiazol-2-amine	1.56	2	MIPS-0000581

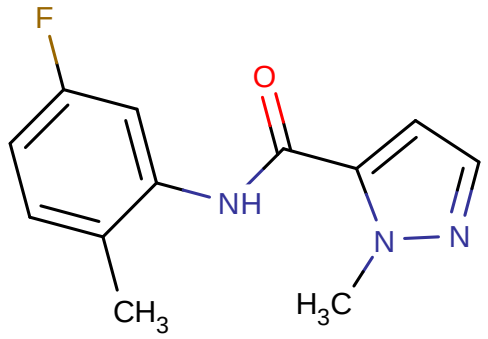
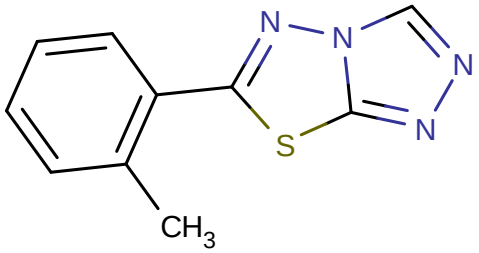
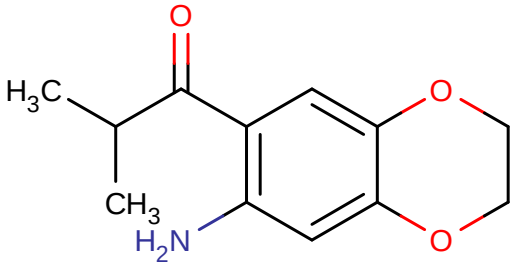
Structure	Plate Name	Plate Well	Mol Weight	Formula	IUPAC Name	LogP	Rotatable bonds	Molecule Name
	2011_Life4	H2	238.24	C14H10N2O2	1-phenyl-1H-1,3-benzodiazole-5-carboxylic acid	0.92	2	MIPS-0000619
	2011_Chembridge 2	F9	220.29	C11H12N2OS	4-(4-methoxyphenyl)-5-methyl-1,3-thiazol-2-amine	3.01	2	MIPS-0000620
	2011_Chembridge 2	H2	207.25	C9H9N3OS	3-amino-6-methylthieno[2,3-b]pyridine-2-carboxamide	0.93	1	MIPS-0000626

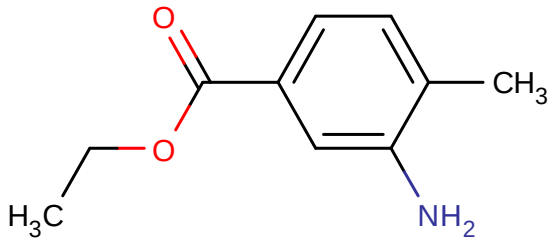
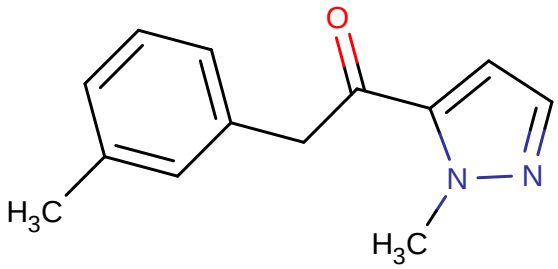
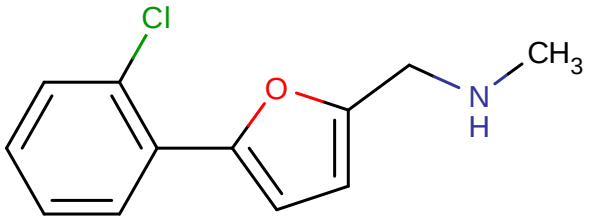
Structure	Plate Name	Plate Well	Mol Weight	Formula	IUPAC Name	LogP	Rotatable bonds	Molecule Name
	2011_ Chembridge 2	H5	199.27	C9H13NO2S	4-propylbenzene-1-sulfonamide	1.98	3	MIPS-0000633
	2011_ Chembridge 2	G7	226.27	C14H14N2O	3-amino-N-(4-methylphenyl)benzamide	2.75	2	MIPS-0000639
	2011_ Life5	A11	215.25	C12H13N3O	N-benzyl-1-methyl-1H-pyrazole-5-carboxamide	1.12	3	MIPS-0000647

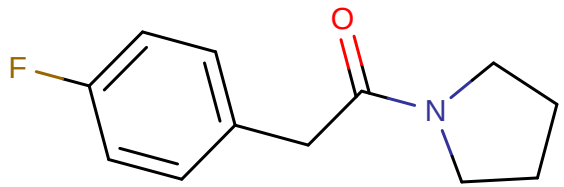
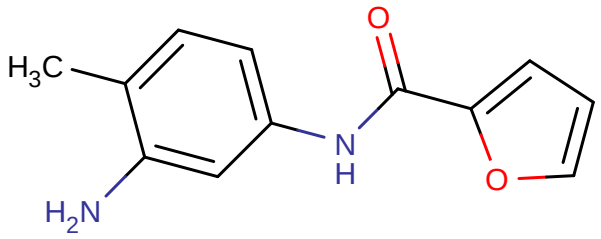
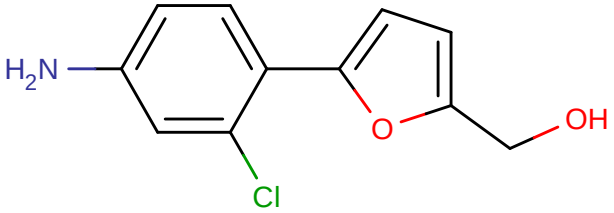
Structure	Plate Name	Plate Well	Mol Weight	Formula	IUPAC Name	LogP	Rotatable bonds	Molecule Name
	2011_ Life5	B9	174.20	C10H10N2O	3-methyl-4-phenyl-1,2-oxazol-5-amine	1.38	1	MIPS-0000650
	2011_ Chembridge 3	B4	190.27	C10H10N2S	5-methyl-4-phenyl-1,3-thiazol-2-amine	3.17	1	MIPS-0000675
	2011_ Chembridge 3	B2	203.19	C11H9NO3	N-(4-hydroxyphenyl)furan-2-carboxamide	1.82	2	MIPS-0000678

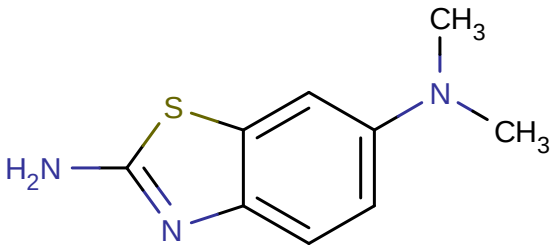
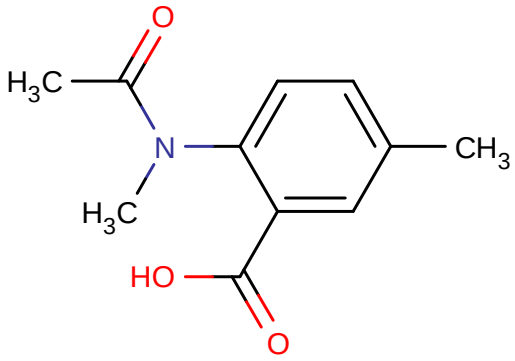
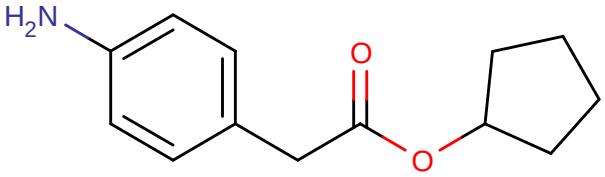
Structure	Plate Name	Plate Well	Mol Weight	Formula	IUPAC Name	LogP	Rotatable bonds	Molecule Name
	2011_ Chembridge 3	E5	202.21	C11H10N2O2	N-(3-aminophenyl)furan-2-carboxamide	1.30	2	MIPS-0000728
	2011_ Life5	H10	208.24	C8H8N4OS	1-methyl-N-(1,3-thiazol-2-yl)-1H-pyrazole-5-carboxamide	0.76	2	MIPS-0000776
	2011_ Life6	D1	231.70	C9H10ClNO2S	N-(3-chlorophenyl)cyclopropanesulfonamide	1.72	2	MIPS-0000797

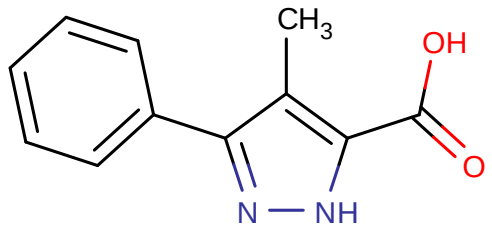
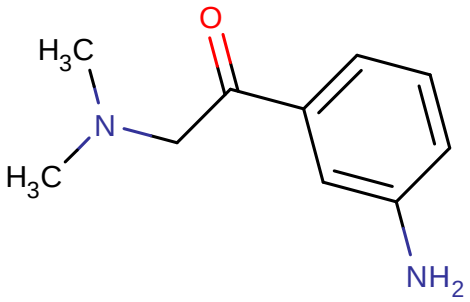
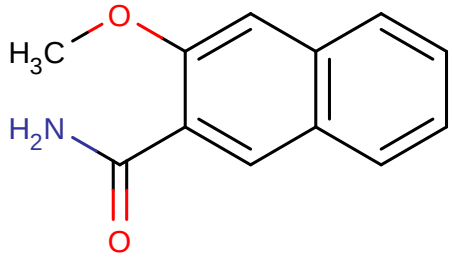
Structure	Plate Name	Plate Well	Mol Weight	Formula	IUPAC Name	LogP	Rotatable bonds	Molecule Name
	2011_ Chembridge 4	C4	164.14	C7H5FN4	5-(4-fluorophenyl)-1H-1,2,3,4-tetrazole	1.40	1	MIPS-0000831
	2011_ Chembridge 5	A2	220.23	C9H5FN4S	6-(3-fluorophenyl)-[1,2,4]triazolo[3,4-b][1,3,4]thiadiazole	1.70	1	MIPS-0000925
	2011_ Chembridge 5	A11	169.65	C9H12ClN	3-(4-chlorophenyl)propan-1-amine	2.44	3	MIPS-0000946

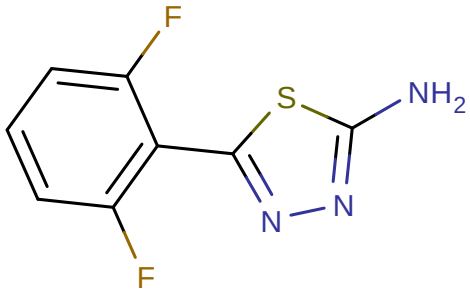
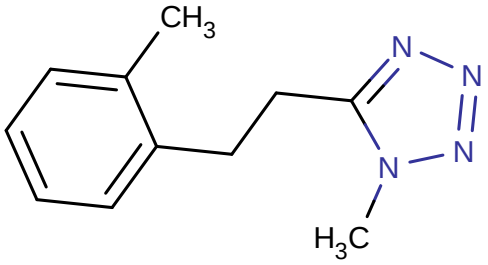
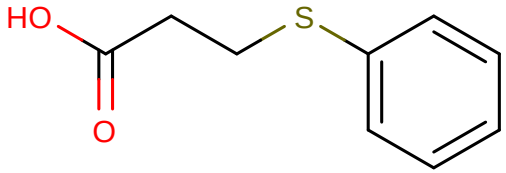
Structure	Plate Name	Plate Well	Mol Weight	Formula	IUPAC Name	LogP	Rotatable bonds	Molecule Name
	2011_ Chembridge 9	A8	233.24	C12H12FN3O	N-(5-fluoro-2-methylphenyl)-1-methyl-1H-pyrazole-5-carboxamide	2.07	2	MIPS-0000979
	2011_ Chembridge 5	D2	216.26	C10H8N4S	6-(2-methylphenyl)-[1,2,4]triazolo[3,4-b][1,3,4]thiadiazole	2.03	1	MIPS-0000984
	2011_ Chembridge 9	B7	221.25	C12H15NO3	1-(7-amino-2,3-dihydro-1,4-benzodioxin-6-yl)-2-methylpropan-1-one	2.11	2	MIPS-0000990

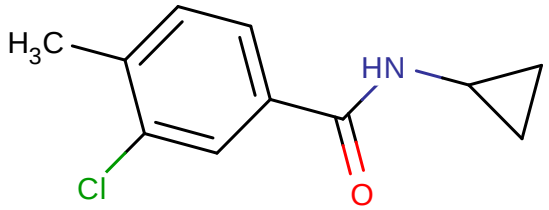
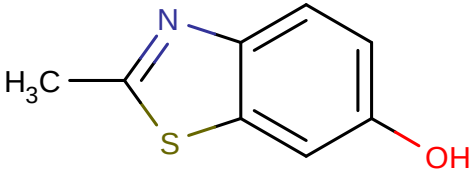
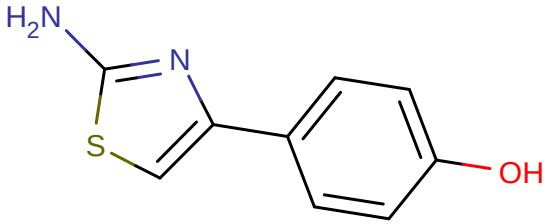
Structure	Plate Name	Plate Well	Mol Weight	Formula	IUPAC Name	LogP	Rotatable bonds	Molecule Name
	2011_ Chembridge 5	C11	179.22	C10H13NO2	ethyl 3-amino-4-methylbenzoate	2.02	3	MIPS-0000995
	2011_ Chembridge 5	E3	215.25	C12H13N3O	1-methyl-N-(3-methylphenyl)-1H-pyrazole-5-carboxamide	1.93	2	MIPS-0001014
	2011_ Chembridge 5	E4	221.68	C12H12ClNO	{[5-(2-chlorophenyl)furan-2-yl]methyl}(methyl)amine	2.76	3	MIPS-0001017

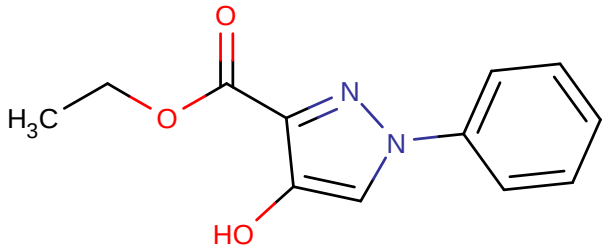
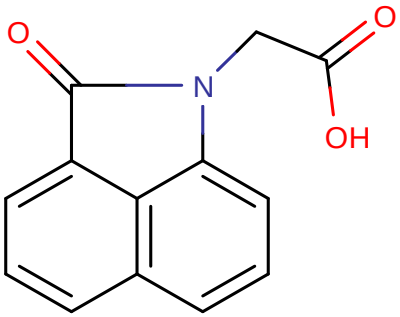
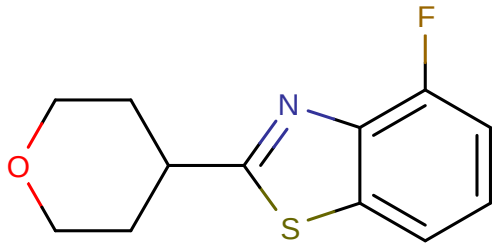
Structure	Plate Name	Plate Well	Mol Weight	Formula	IUPAC Name	LogP	Rotatable bonds	Molecule Name
	2011_ Chembridge 9	F3	207.24	C12H14FNO	2-(4-fluorophenyl)-1-(pyrrolidin-1-yl)ethan-1-one	1.80	2	MIPS-0001049
	2011_ Chembridge 5	H2	216.24	C12H12N2O2	N-(3-amino-4-methylphenyl)furan-2-carboxamide	1.81	2	MIPS-0001058
	2011_ Chembridge 9	H9	223.66	C11H10ClNO2	[5-(4-amino-2-chlorophenyl)furan-2-yl]methanol	1.61	2	MIPS-0001085

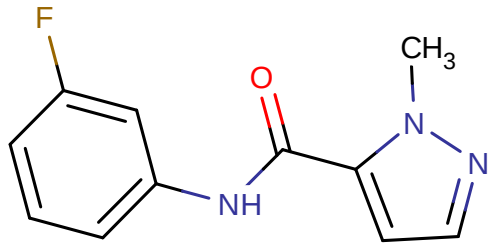
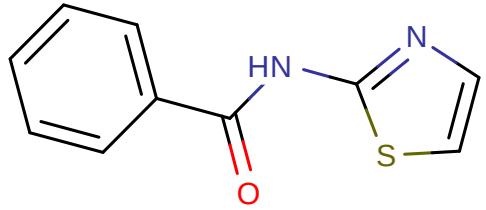
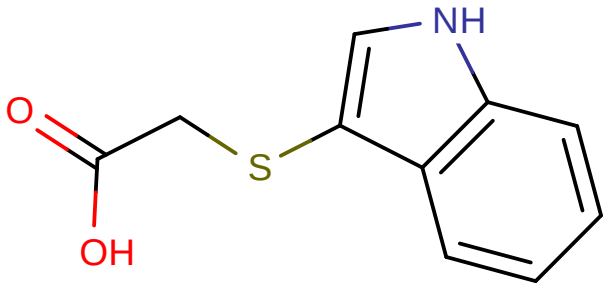
Structure	Plate Name	Plate Well	Mol Weight	Formula	IUPAC Name	LogP	Rotatable bonds	Molecule Name
	2011_ Chembridge 10	A6	193.27	C9H11N3S	6-N,6-N-dimethyl-1,3-benzothiazole-2,6-diamine	2.08	1	MIPS-0001105
	2011_ Chembridge 10	C11	207.23	C11H13NO3	5-methyl-2-(N-methylacetamido)benzoic acid	1.25	2	MIPS-0001157
	2011_ Chembridge 10	D1	205.25	C12H15NO2	cyclopentyl 4-aminobenzoate	2.50	3	MIPS-0001161

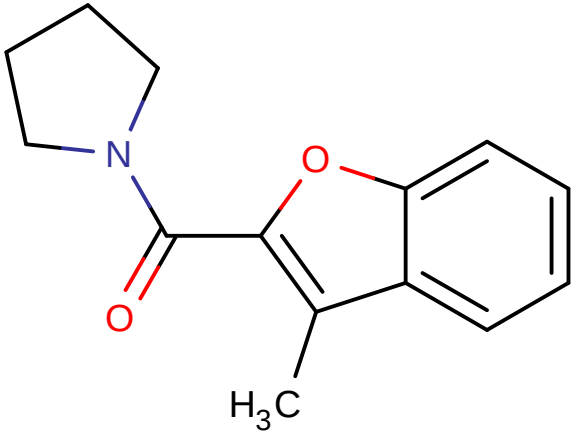
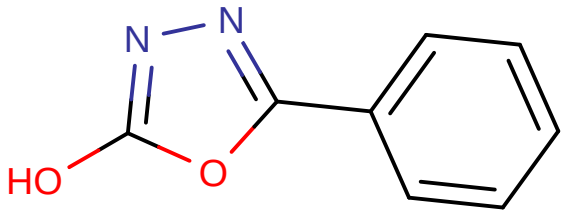
Structure	Plate Name	Plate Well	Mol Weight	Formula	IUPAC Name	LogP	Rotatable bonds	Molecule Name
	2011_ Chembridge 10	E3	202.21	C11H10N2O2	4-methyl-3-phenyl-1H-pyrazole-5-carboxylic acid	2.40	2	MIPS-0001179
	2011_ Chembridge 6	G5	178.23	C10H14N2O	3-amino-N-(propan-2-yl)benzamide	0.99	2	MIPS-0001192
	2011_ Chembridge 7	B2	201.22	C12H11NO2	3-methoxynaphthalene-2-carboxamide	1.66	2	MIPS-0001244

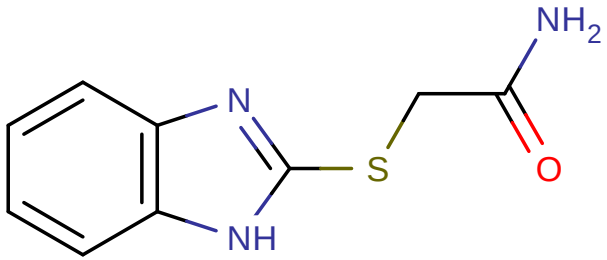
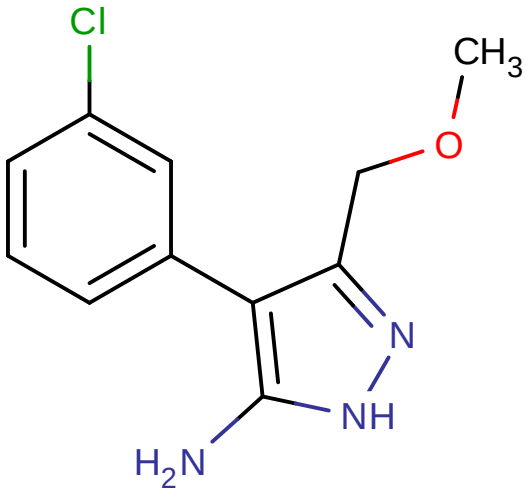
Structure	Plate Name	Plate Well	Mol Weight	Formula	IUPAC Name	LogP	Rotatable bonds	Molecule Name
	2011_ Chembridge 14	H4	213.21	C8H5F2N3S	5-(2,6-difluorophenyl)-1,3,4-thiadiazol-2-amine	1.91	1	MIPS-0001260
	2011_ Chembridge 14	E5	220.29	C10H12N4S	1-methyl-5-{[(2-methylphenyl)methyl]sulfanyl}-1H-1,2,3,4-tetrazole	2.80	3	MIPS-0001303
	2011_ Life1	A11	182.24	C9H10O2S	3-(phenylsulfanyl)propanoic acid	2.21	4	MIPS-0001305

Structure	Plate Name	Plate Well	Mol Weight	Formula	IUPAC Name	LogP	Rotatable bonds	Molecule Name
	2011_ Chembridge 14	D5	209.67	C11H12ClNO	3-chloro-N-cyclopropyl-4-methylbenzamide	2.63	2	MIPS-0001314
	2011_ Life1	F5	165.21	C8H7NOS	2-methyl-1,3-benzothiazol-6-ol	1.93	0	MIPS-0001366
	2011_ Life1	G4	192.24	C9H8N2OS	4-(2-amino-1,3-thiazol-4-yl)phenol	2.22	1	MIPS-0001371

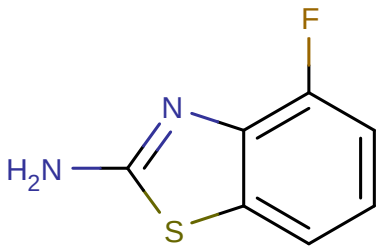
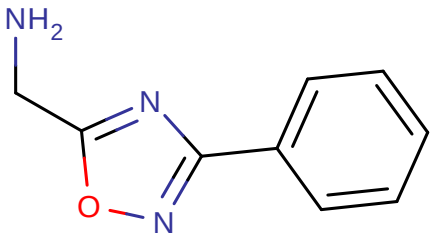
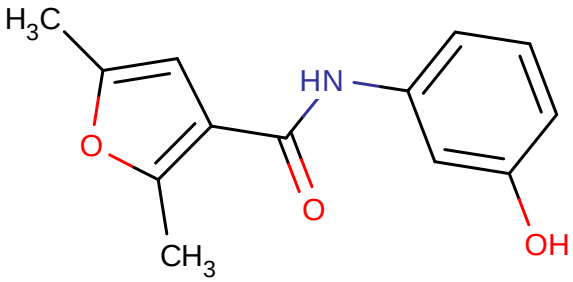
Structure	Plate Name	Plate Well	Mol Weight	Formula	IUPAC Name	LogP	Rotatable bonds	Molecule Name
	2011_ Life1	G3	232.24	C12H12N2O3	ethyl 4-hydroxy-1-phenyl-1H-pyrazole-3-carboxylate	2.95	4	MIPS-0001377
	2011_ Life1	G8	227.22	C13H9NO3	2-{3-oxo-2-azatricyclo[6.3.1.0^{4},^{12}]dodeca-1(11),4(12),5,7,9-pentaen-2-yl}acetic acid	1.42	2	MIPS-0001384
	2011_ Life1	H7	238.28	C11H11FN2OS	4-fluoro-2-(morpholin-4-yl)-1,3-benzothiazole	2.83	1	MIPS-0001391

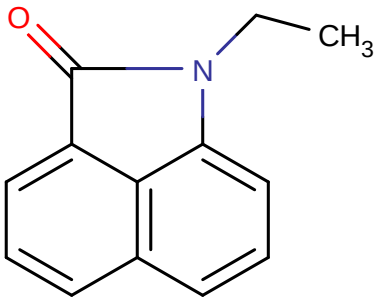
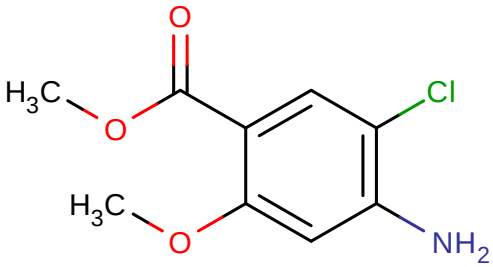
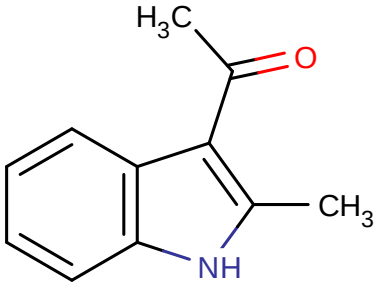
Structure	Plate Name	Plate Well	Mol Weight	Formula	IUPAC Name	LogP	Rotatable bonds	Molecule Name
	2011_Chembridge 11	C7	219.22	C11H10FN3O	N-(3-fluorophenyl)-1-methyl-1H-pyrazole-5-carboxamide	1.56	2	MIPS-0001396
	2011_Life1	H8	204.25	C10H8N2OS	N-(1,3-thiazol-2-yl)benzamide	2.41	2	MIPS-0001397
	2011_Life1	H10	207.25	C10H9NO2S	2-(1H-indol-3-ylsulfanyl)acetic acid	1.87	3	MIPS-0001400

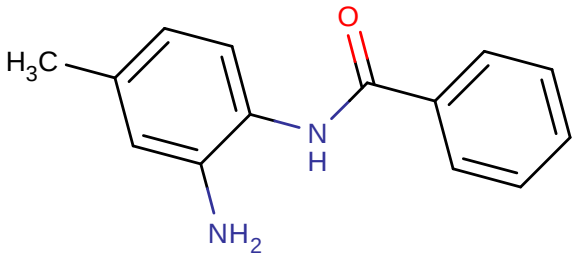
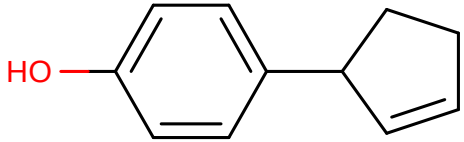
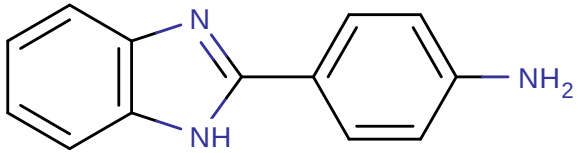
Structure	Plate Name	Plate Well	Mol Weight	Formula	IUPAC Name	LogP	Rotatable bonds	Molecule Name
	2011_ Chembridge 11	D10	229.27	C14H15NO2	1-(3-methyl-1-benzofuran-2-carbonyl)pyrrolidine	2.27	1	MIPS-0001403
	2011_ Life2	A9	162.15	C8H6N2O2	5-phenyl-1,3,4-oxadiazol-2-ol	1.36	1	MIPS-0001417

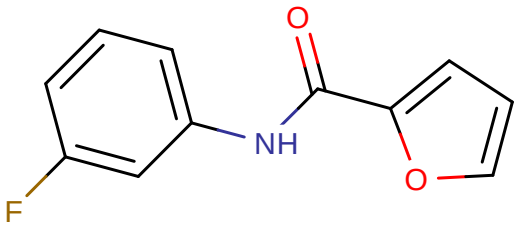
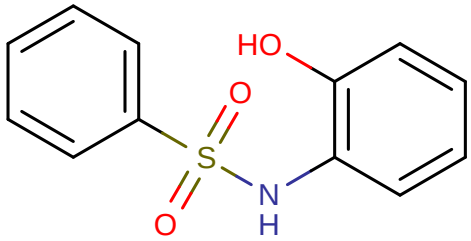
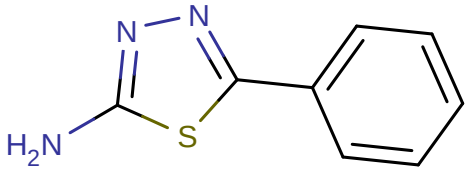
Structure	Plate Name	Plate Well	Mol Weight	Formula	IUPAC Name	LogP	Rotatable bonds	Molecule Name
	2011_ Chembridge 11	D12	207.25	C9H9N3OS	2-(1H-1,3-benzodiazol-2-ylsulfanyl)acetamide	0.93	3	MIPS-0001424
	2011_ Chembridge 11	G1	237.69	C11H12ClN3O	4-(3-chlorophenyl)-3-(methoxymethyl)-1H-pyrazol-5-amine	1.75	3	MIPS-0001448

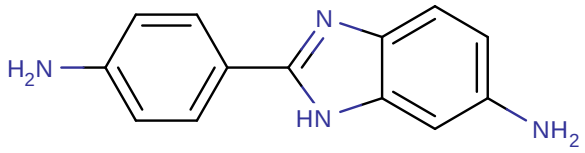
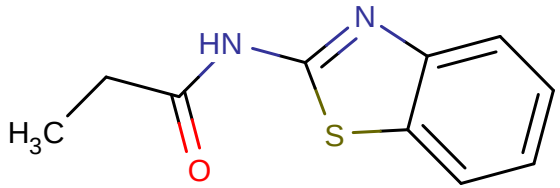
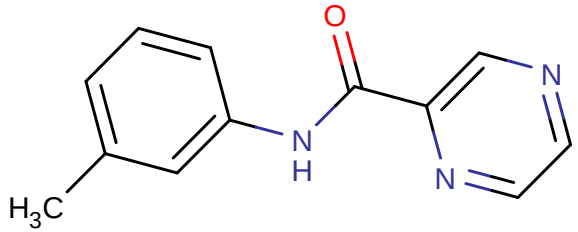
STD rank 3

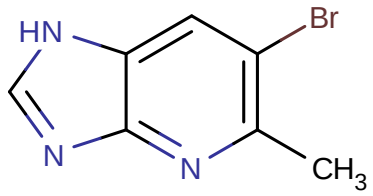
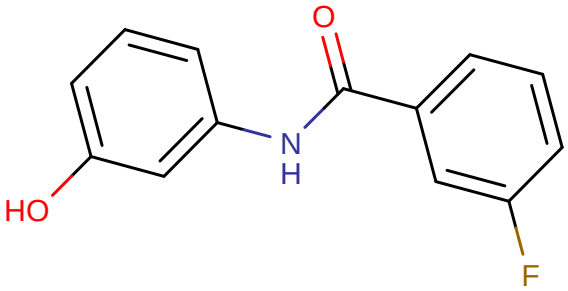
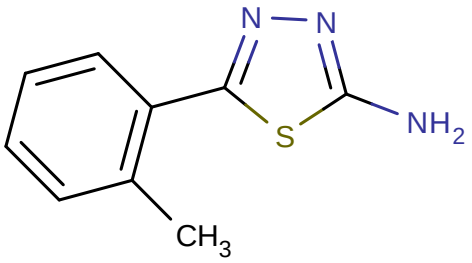
Structure	Plate name	Plate well	Mol weight	Formula	IUPAC name	LogP	Rotatable bonds	Molecule name
	2011_Life2	D12	168.19	C7H5FN2S	4-fluoro-1,3-benzothiazol-2-amine	2.11	0	MIPS-0001463
	2011_Life3	B12	175.19	C9H9N3O	(3-phenyl-1,2,4-oxadiazol-5-yl)methanamine	1.45	2	MIPS-0000415
	2011_Chembridge3	D6	231.25	C13H13NO3	N-(3-hydroxyphenyl)-2,5-dimethylfuran-3-carboxamide	2.30	2	MIPS-0000713

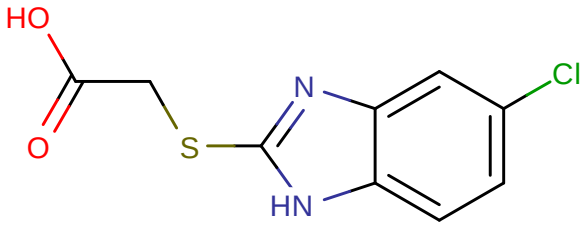
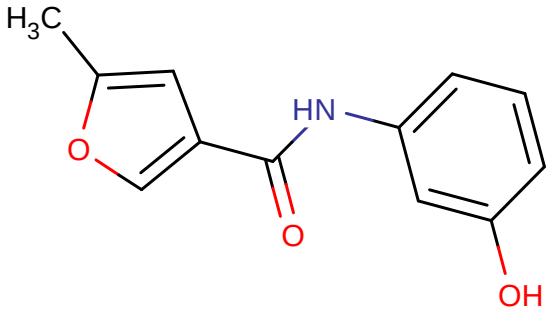
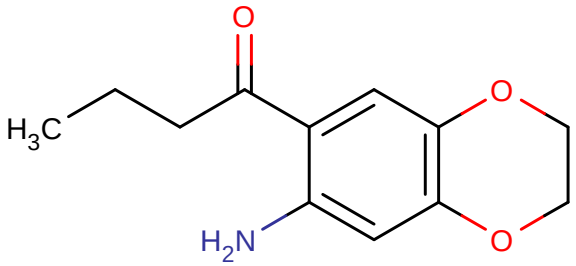
Structure	Plate name	Plate well	Mol weight	Formula	IUPAC name	LogP	Rotatable bonds	Molecule name
	2011_Life1	C6	197.23	C13H11NO	2-ethyl-2-azatricyclo[6.3.1.$0^{4,12}$]dodeca-1(11),4(12),5,7,9-pentaen-3-one	2.30	1	MIPS-0001322
	2011_Chembridge4	D4	215.63	C9H10ClNO3	methyl 4-amino-5-chloro-2-methoxybenzoate	1.59	3	MIPS-0000861
	2011_Life2	A7	173.21	C11H11NO	1-(2-methyl-1H-indol-3-yl)ethan-1-one	1.83	1	MIPS-0001405

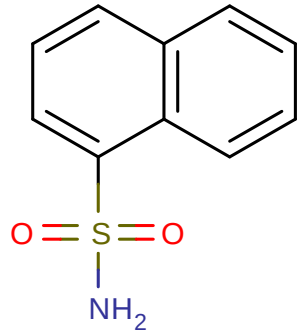
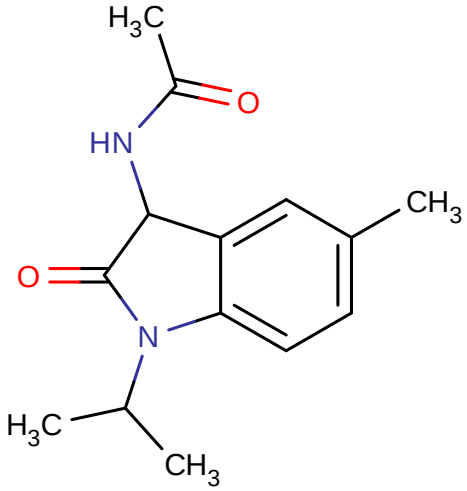
Structure	Plate name	Plate well	Mol weight	Formula	IUPAC name	LogP	Rotatable bonds	Molecule name
	2011_Chembridge1	E10	226.27	C14H14N2O	N-(2-amino-4-methylphenyl)benzamide	2.75	2	MIPS-0000443
	2011_Life1	B8	160.21	C11H12O	4-(cyclopent-2-en-1-yl)phenol	2.98	1	MIPS-0001311
	2011_Chembridge1	C12	209.25	C13H11N3	4-(1H-1,3-benzodiazol-2-yl)aniline	2.46	1	MIPS-0000400

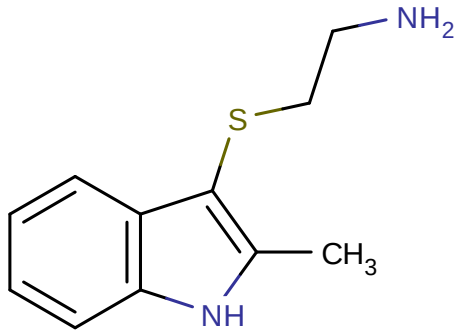
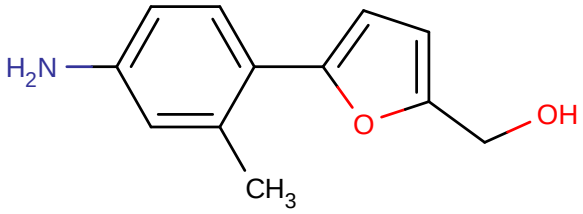
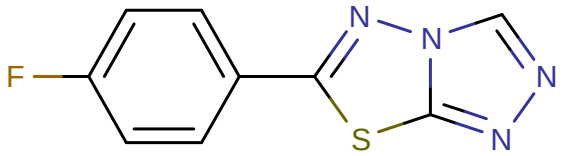
Structure	Plate name	Plate well	Mol weight	Formula	IUPAC name	LogP	Rotatable bonds	Molecule name
	2011_Chembridge2	A7	205.19	C11H8FNO2	N-(3-fluorophenyl)furan-2-carboxamide	2.27	2	MIPS-0000498
	2011_Chembridge2	F2	249.29	C12H11NO3S	N-(2-hydroxyphenyl)benzenesulfonamide	2.81	2	MIPS-0000589
	2011_Chembridge2	F4	177.23	C8H7N3S	5-phenyl-1,3,4-thiadiazol-2-amine	1.63	1	MIPS-0000603

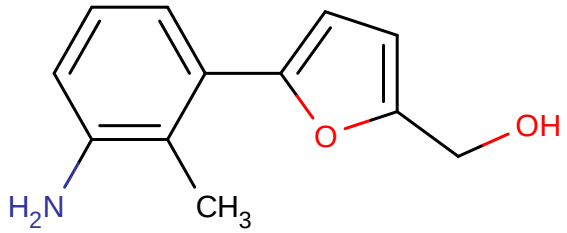
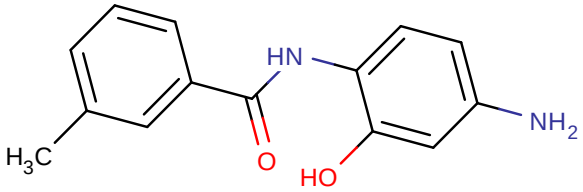
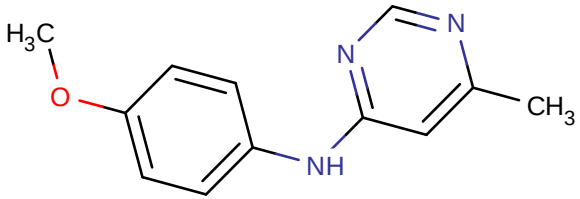
Structure	Plate name	Plate well	Mol weight	Formula	IUPAC name	LogP	Rotatable bonds	Molecule name
	2011_Chembridge2	G4	224.26	C13H12N4	2-(4-aminophenyl)-1H-1,3-benzodiazol-6-amine	1.63	1	MIPS-0000615
	2011_Chembridge3	G3	210.30	C10H14N2OS	N-(4,5,6,7-tetrahydro-1,3-benzothiazol-2-yl)propanamide	2.71	2	MIPS-0000752
	2011_Chembridge3	E4	213.24	C12H11N3O	N-(3-methylphenyl)pyrazine-2-carboxamide	1.53	2	MIPS-0000725

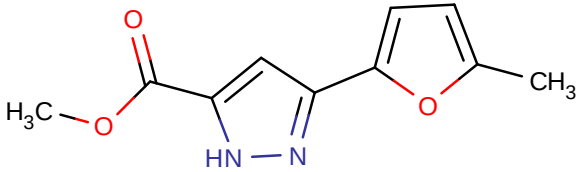
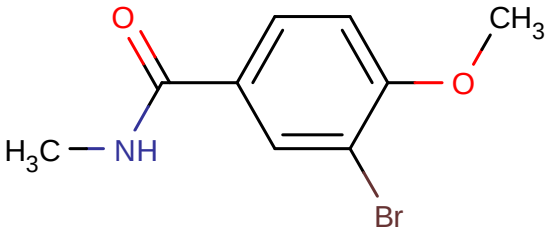
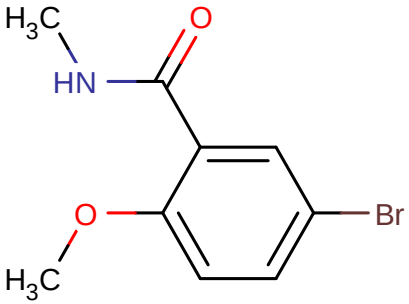
Structure	Plate name	Plate well	Mol weight	Formula	IUPAC name	LogP	Rotatable bonds	Molecule name
	2011_Life2	D11	212.05	C7H6BrN3	6-bromo-5-methyl-1H-imidazo[4,5-b]pyridine	1.27	0	MIPS-0001462
	2011_Chembridge4	F6	231.22	C13H10FNO2	3-fluoro-N-(3-hydroxyphenyl)benzamide	2.90	2	MIPS-0000890
	2011_Chembridge4	H6	191.25	C9H9N3S	5-(2-methylphenyl)-1,3,4-thiadiazol-2-amine	2.14	1	MIPS-0000914

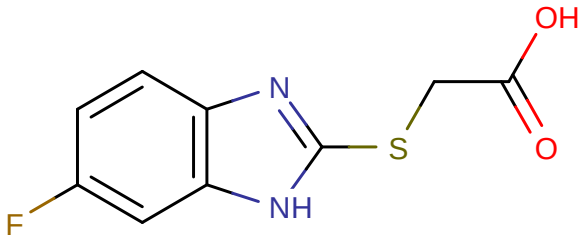
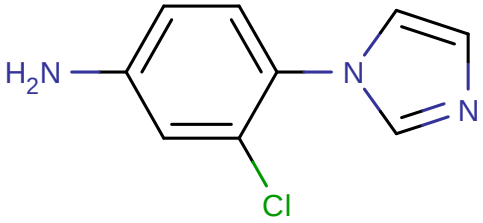
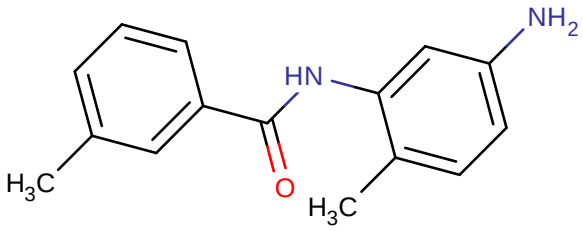
Structure	Plate name	Plate well	Mol weight	Formula	IUPAC name	LogP	Rotatable bonds	Molecule name
	2011_Chembridge14	G6	242.68	C9H7ClN2O2S	2-[(5-chloro-1H-1,3-benzodiazol-2-yl)sulfanyl]acetic acid	1.54	3	MIPS-0001282
	2011_Chembridge4	E7	217.22	C12H11NO3	N-(3-hydroxyphenyl)-5-methylfuran-3-carboxamide	2.10	2	MIPS-0000873
	2011_Chembridge9	A4	221.25	C12H15NO3	1-(7-amino-2,3-dihydro-1,4-benzodioxin-6-yl)butan-1-one	2.01	3	MIPS-0000968

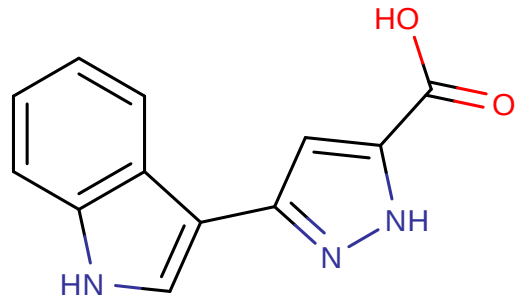
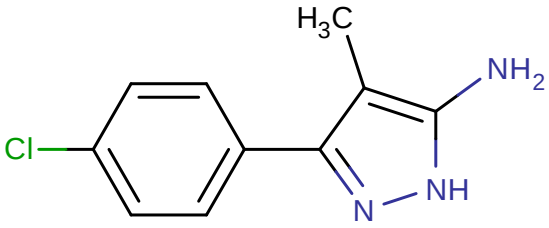
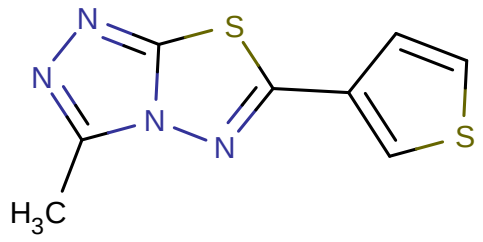
Structure	Plate name	Plate well	Mol weight	Formula	IUPAC name	LogP	Rotatable bonds	Molecule name
	2011_Chembridge4	A11	207.25	C10H9NO2S	naphthalene-1-sulfonamide	1.57	1	MIPS-0000795
	2011_Chembridge9	G5	246.30	C14H18N2O2	N-[5-methyl-2-oxo-1-(propan-2-yl)-2,3-dihydro-1H-indol-3-yl]acetamide	1.22	2	MIPS-0001071

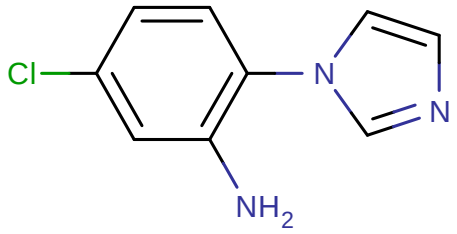
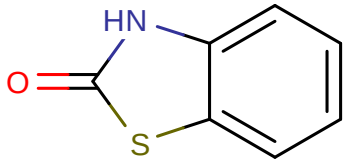
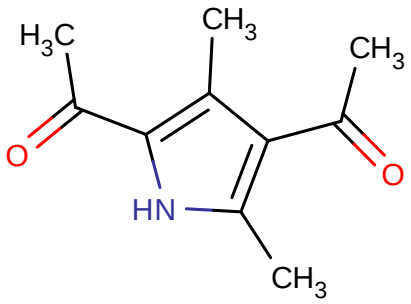
Structure	Plate name	Plate well	Mol weight	Formula	IUPAC name	LogP	Rotatable bonds	Molecule name
	2011_Life2	F5	206.31	C11H14N2S	3-[(2-aminoethyl)sulfanyl]-2-methyl-1H-indole	1.84	3	MIPS-0001484
	2011_Chembridge9	G9	203.24	C12H13NO2	[5-(4-amino-2-methylphenyl)furan-2-yl]methanol	1.52	2	MIPS-0001074
	2011_Chembridge5	C1	220.23	C9H5FN4S	6-(4-fluorophenyl)-[1,2,4]triazolo[3,4-b][1,3,4]thiadiazole	1.70	1	MIPS-0000960

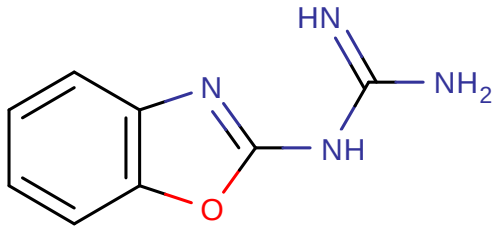
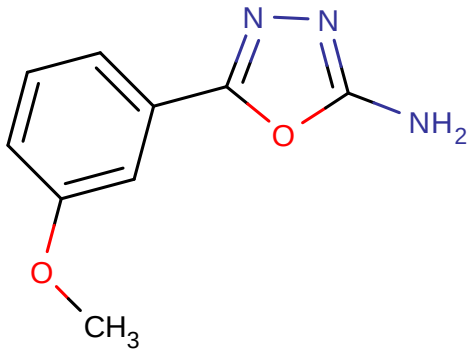
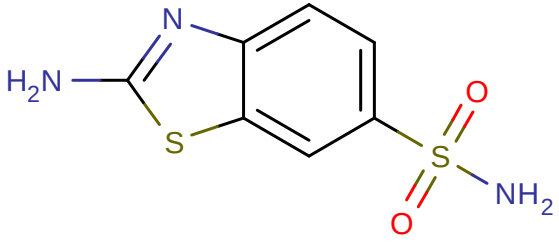
Structure	Plate name	Plate well	Mol weight	Formula	IUPAC name	LogP	Rotatable bonds	Molecule name
	2011_Chembridge9	G10	203.24	C12H13NO2	[5-(3-amino-2-methylphenyl)furan-2-yl]methanol	1.52	2	MIPS-0001075
	2011_Chembridge9	C12	242.27	C14H14N2O2	N-(4-amino-2-hydroxyphenyl)-3-methylbenzamide	2.45	2	MIPS-0001011
	2011_Life1	F3	215.25	C12H13N3O	N-(4-methoxyphenyl)-6-methylpyrimidin-4-amine	2.06	3	MIPS-0001363

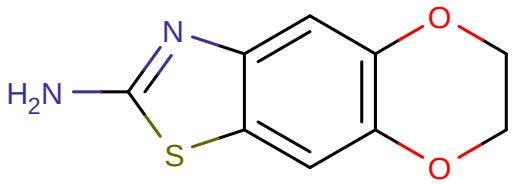
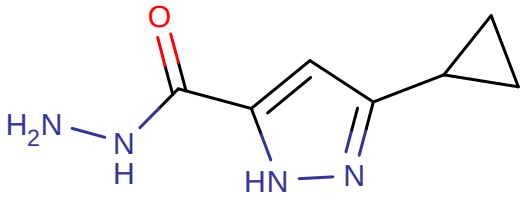
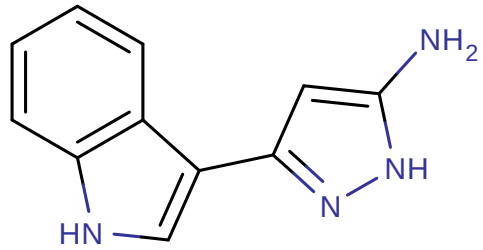
Structure	Plate name	Plate well	Mol weight	Formula	IUPAC name	LogP	Rotatable bonds	Molecule name
	2011_Life5	C6	206.20	C10H10N2O3	methyl 3-(5-methylfuran-2-yl)-1H-pyrazole-5-carboxylate	1.49	3	MIPS-0000682
	2011_Chembridge6	G9	244.09	C9H10BrNO2	3-bromo-4-methoxy-N-methylbenzamide	1.66	2	MIPS-0001202
	2011_Chembridge6	E10	244.09	C9H10BrNO2	5-bromo-2-methoxy-N-methylbenzamide	1.66	2	MIPS-0001165

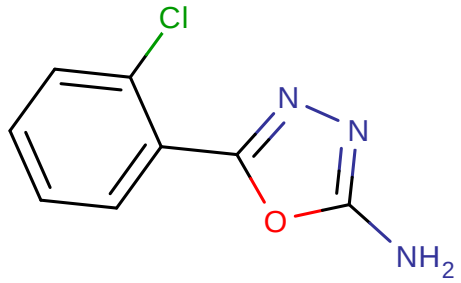
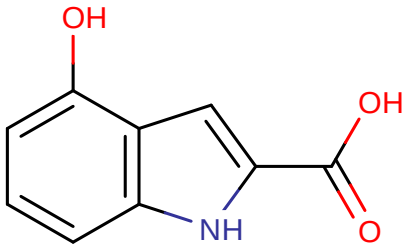
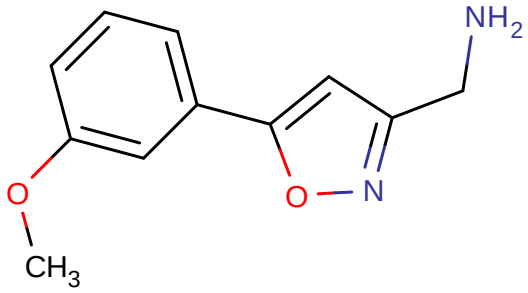
Structure	Plate name	Plate well	Mol weight	Formula	IUPAC name	LogP	Rotatable bonds	Molecule name
	2011_Chembridge14	C2	226.23	C9H7FN2O2S	2-[(6-fluoro-1H-1,3-benzodiazol-2-yl)sulfanyl]acetic acid	0.92	3	MIPS-0001327
	2011_Chembridge14	D2	193.63	C9H8ClN3	3-chloro-4-(1H-imidazol-1-yl)aniline	1.00	1	MIPS-0001307
	2011_Chembridge10	F10	240.30	C15H16N2O	N-(5-amino-2-methylphenyl)-3-methylbenzamide	3.26	2	MIPS-0001211

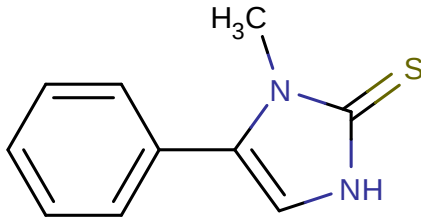
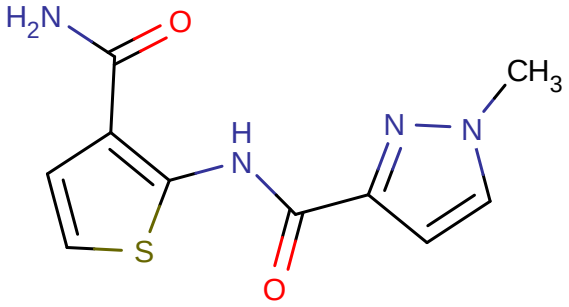
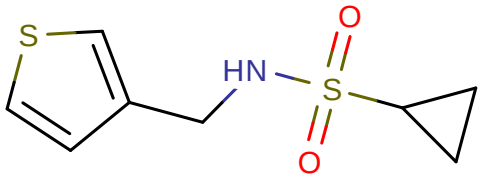
Structure	Plate name	Plate well	Mol weight	Formula	IUPAC name	LogP	Rotatable bonds	Molecule name
	2011_Life5	C9	227.22	C12H9N3O2	3-(1H-indol-3-yl)-1H-pyrazole-5-carboxylic acid	1.99	2	MIPS-0000676
	2011_Chembridge6	F3	207.66	C10H10ClN3	3-(4-chlorophenyl)-4-methyl-1H-pyrazol-5-amine	2.69	1	MIPS-0001176
	2011_Chembridge10	H8	222.29	C8H6N4S2	3-methyl-6-(thiophen-3-yl)-[1,2,4]triazolo[3,4-b][1,3,4]thiadiazole	1.52	1	MIPS-0001249

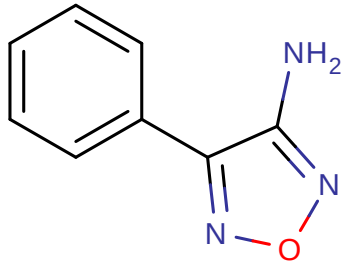
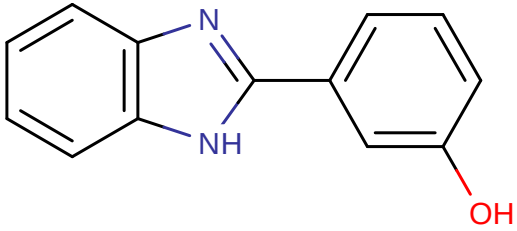
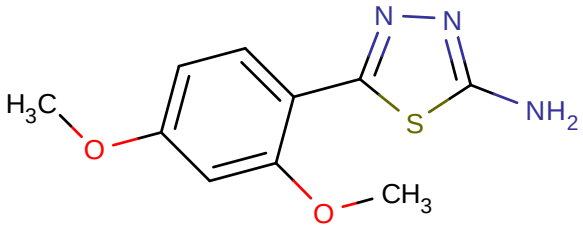
Structure	Plate name	Plate well	Mol weight	Formula	IUPAC name	LogP	Rotatable bonds	Molecule name
	2011_Chembridge14	D6	193.63	C9H8ClN3	5-chloro-2-(1H-imidazol-1-yl)aniline	1.00	1	MIPS-0001308
	2011_Life1	A12	151.19	C7H5NOS	2,3-dihydro-1,3-benzothiazol-2-one	1.99	0	MIPS-0001294
	2011_Life1	D3	181.19	C9H11NO3	4-acetyl-3,5-dimethyl-1H-pyrrole-2-carboxylic acid	0.90	2	MIPS-0001351

Structure	Plate name	Plate well	Mol weight	Formula	IUPAC name	LogP	Rotatable bonds	Molecule name
	2011_Life3	H1	176.18	C8H8N4O	1-(1,3-benzoxazol-2-yl)guanidine	0.98	1	MIPS-0000471
	2011_Life3	G10	191.19	C9H9N3O2	5-(3-methoxyphenyl)-1,3,4-oxadiazol-2-amine	0.68	2	MIPS-0000477
	2011_Life2	H5	229.28	C7H7N3O2S2	2-amino-1,3-benzothiazole-6-sulfonamide	0.57	1	MIPS-0000361

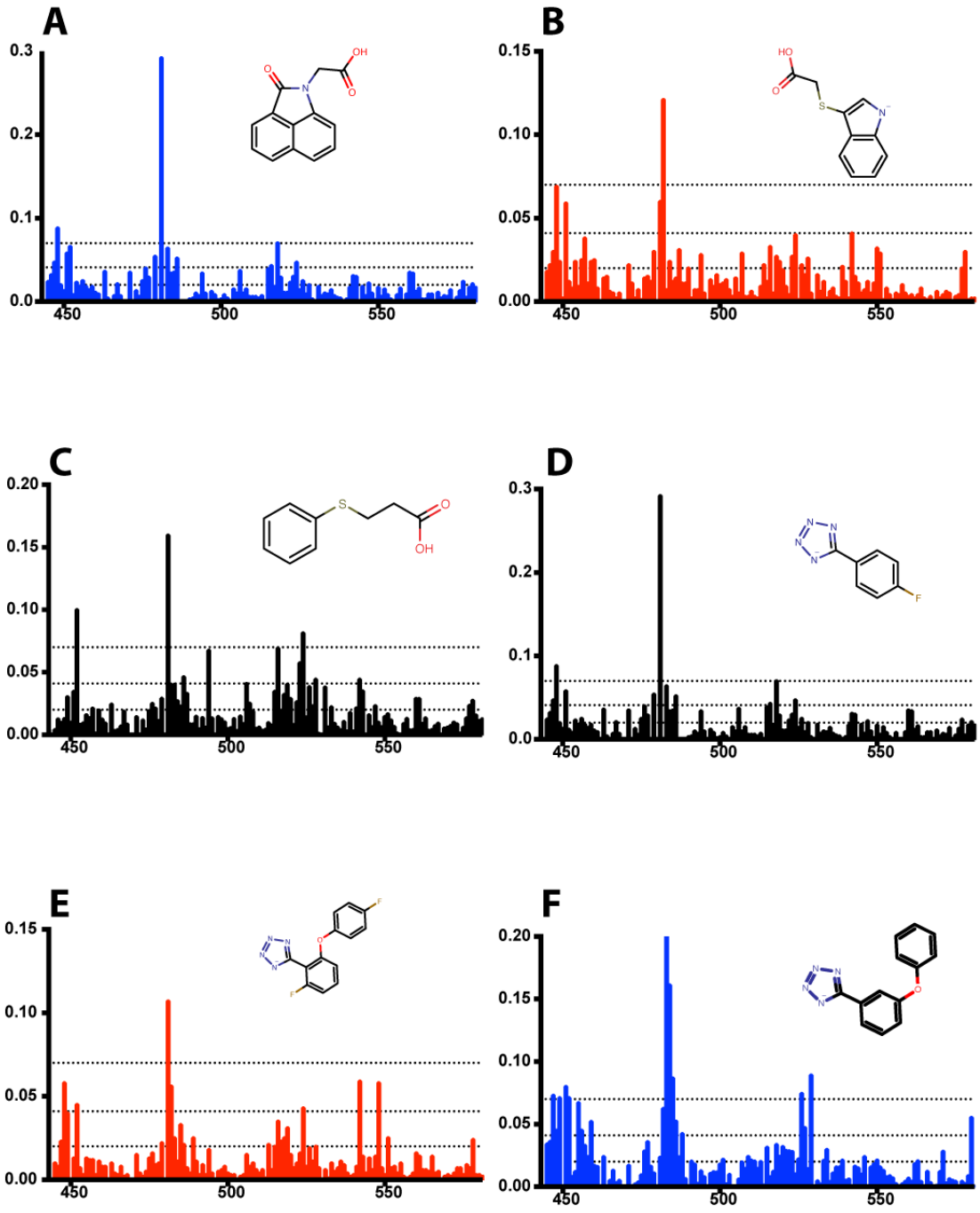
Structure	Plate name	Plate well	Mol weight	Formula	IUPAC name	LogP	Rotatable bonds	Molecule name
	2011_Life2	F12	208.24	C9H8N2O2S	10,13-dioxa-4-thia-6-azatricyclo[7.4.0.$0^{3,7}$]trideca-1(9),2,5,7-tetraen-5-amine	1.48	0	MIPS-0000350
	2011_Life2	G1	166.18	C7H10N4O	3-cyclopropyl-1H-pyrazole-5-carbohydrazide	-0.34	2	MIPS-0000357
	2011_Life3	F9	198.22	C11H10N4	3-(1H-indol-3-yl)-1H-pyrazol-5-amine	1.67	1	MIPS-0000457

Structure	Plate name	Plate well	Mol weight	Formula	IUPAC name	LogP	Rotatable bonds	Molecule name
	2011_Life3	F10	195.61	C8H6ClN3O	5-(2-chlorophenyl)-1,3,4-oxadiazol-2-amine	1.44	1	MIPS-0000458
	2011_Life4	B4	177.16	C9H7NO3	4-hydroxy-1H-indole-2-carboxylic acid	1.35	1	MIPS-0000518
	2011_Life4	B5	204.23	C11H12N2O2	[5-(3-methoxyphenyl)-1,2-oxazol-3-yl]methanamine	0.95	3	MIPS-0000523

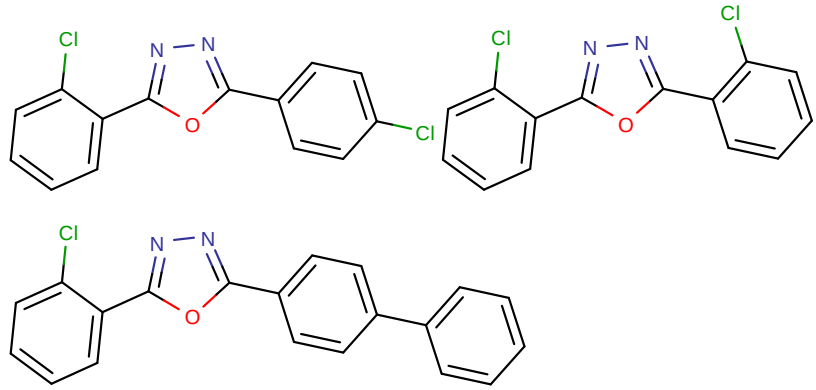
Structure	Plate name	Plate well	Mol weight	Formula	IUPAC name	LogP	Rotatable bonds	Molecule name
	2011_Life5	B2	190.27	C10H10N2S	1-methyl-5-phenyl-2,3-dihydro-1H-imidazole-2-thione	2.04	1	MIPS-0000662
	2011_Life5	B11	250.28	C10H10N4O2S	N-(3-carbamoylthiophen-2-yl)-1-methyl-1H-pyrazole-3-carboxamide	1.33	3	MIPS-0000671
	2011_Life6	B1	217.31	C8H11NO2S2	N-(thiophen-3-ylmethyl)cyclopropanesulfonamide	0.96	3	MIPS-0000791

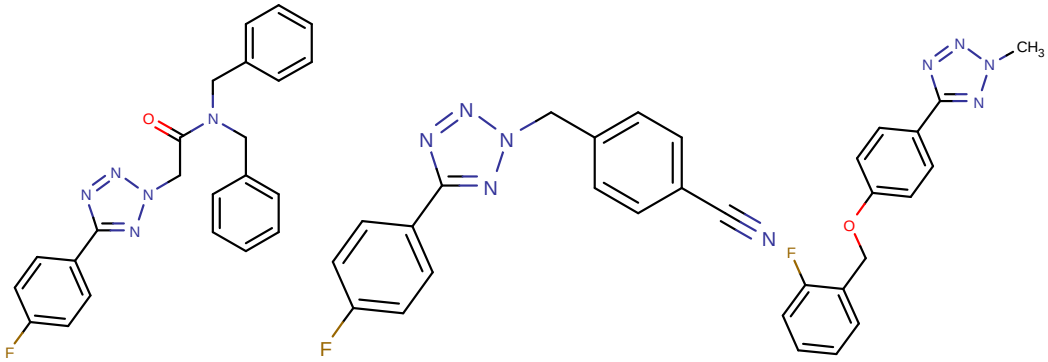
Structure	Plate name	Plate well	Mol weight	Formula	IUPAC name	LogP	Rotatable bonds	Molecule name
	2011_Life1	G5	161.16	C8H7N3O	4-phenyl-1,2,5-oxadiazol-3-amine	1.36	1	MIPS-0001380
	2011_Life2	A3	210.23	C13H10N2O	3-(1H-1,3-benzodiazol-2-yl)phenol	2.98	1	MIPS-0001404
	2011_Life2	A5	237.28	C10H11N3O2S	5-(2,4-dimethoxyphenyl)-1,3,4-thiadiazol-2-amine	1.31	3	MIPS-0001410

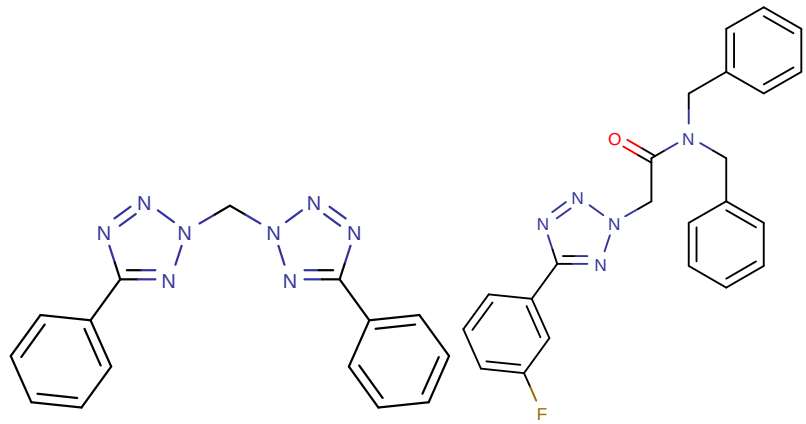
Appendix C. *A CSP plots altered by addition of (A) L1G8, (B) L1H10, (C) L1A11, (D) C4C4, (E) ZINC72447025 and (F) CDS001350 compounds monitored by* ^{15}*N-HSQC.*

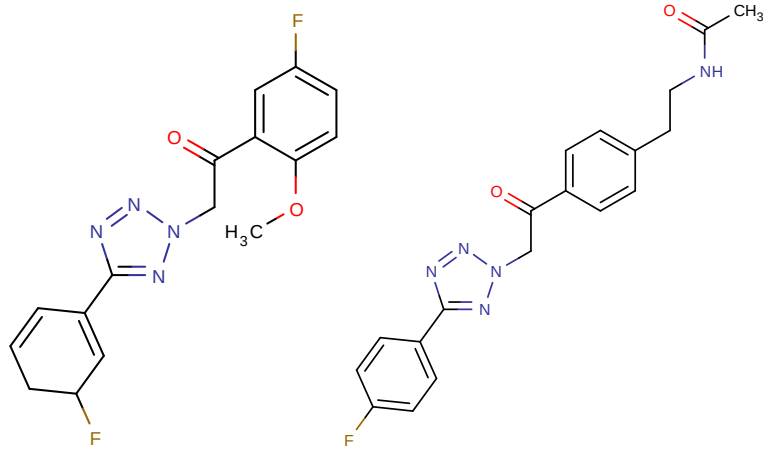


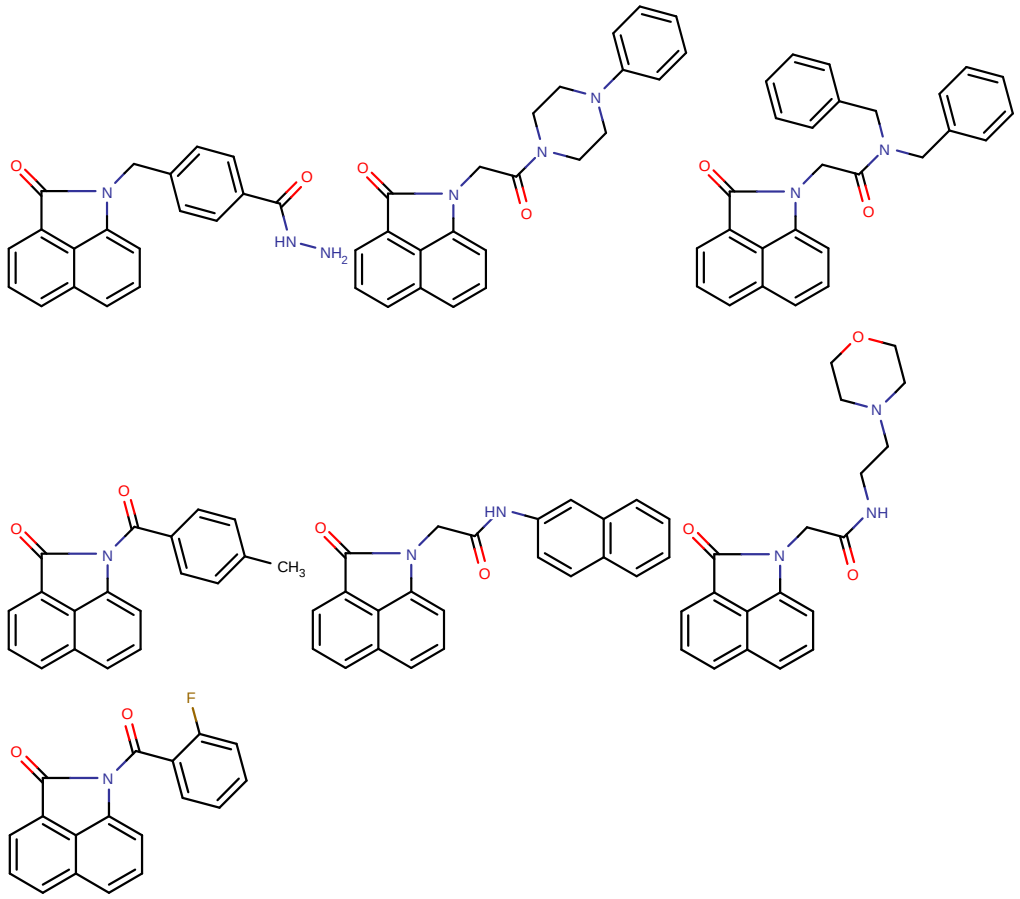
Appendix D. *The list of purchased fragment analogues that did not bind to DnaGC.*

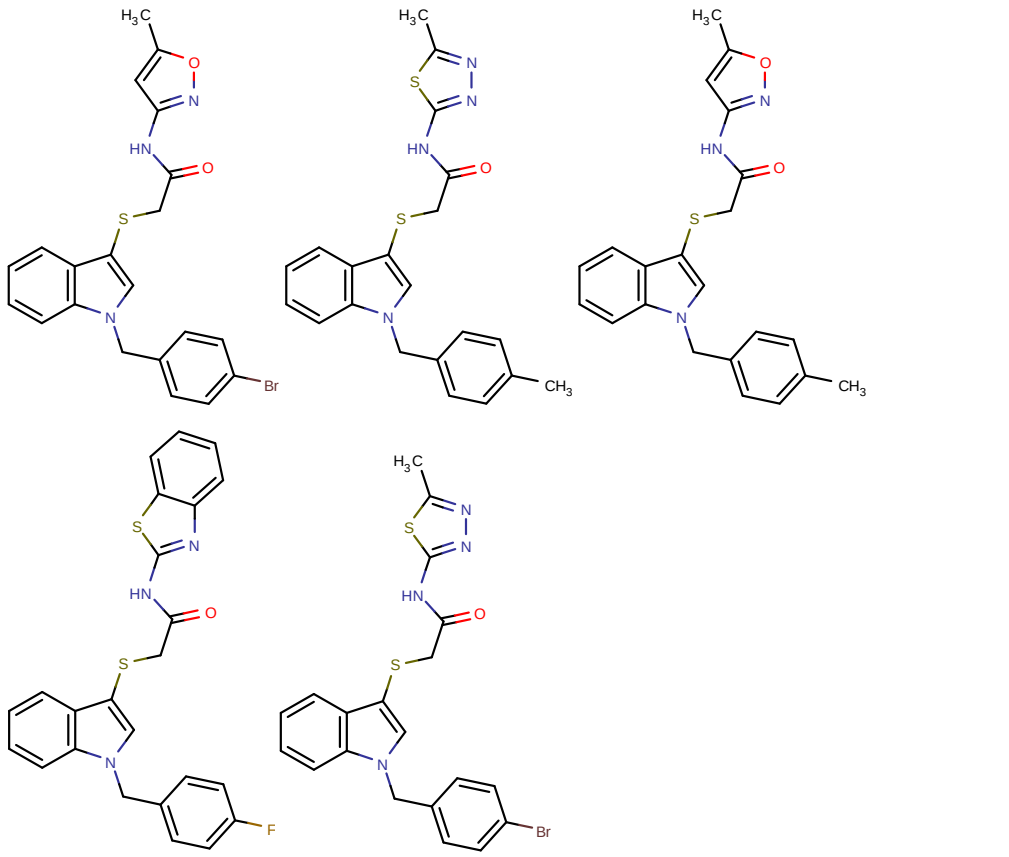




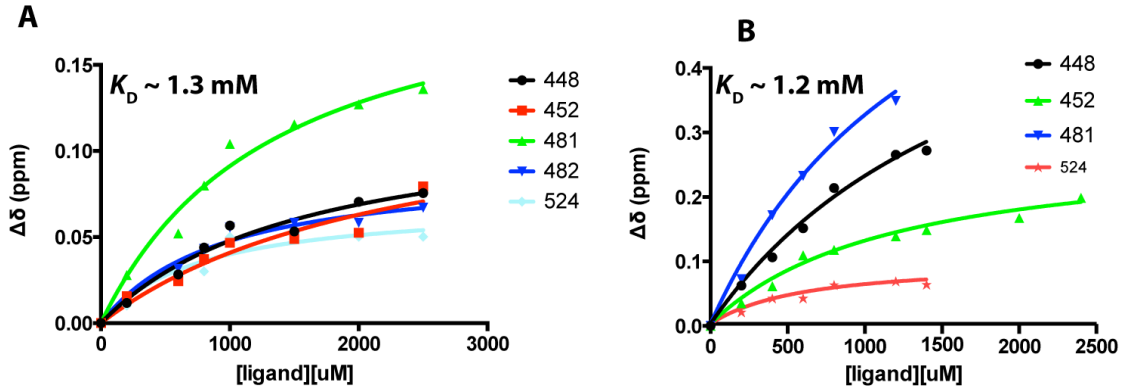








Appendix E. *Binding affinities measured by HSQC titration of (A) ZINC72447025 and (B) CDS00135 compounds. Different curves corresponds to perturbed residues upon addition of fragment.*



Appendix F. *STD-NMR of ZINC72447025 fragment with other SSB-Ct binding partners represented in different colours: black-1D reference spectrum of ZINC72447025, green-STD spectrum of E. coli χ, yellow-A. baumanni χ, dark green-RNAse HI and red-PriA respectively.*